

RESULT	1
ID	W04842 standard: peptide; 15 AA.
AC	W04842;
DT	18-FEB-1997 (first entry)
DE	Self epitope of desmoglein 3, implicated in autoimmune disease
KW	Tolerisation; self-epitope; antigen; autoimmune disease;
KW	autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW	pemphigus vulgaris; desmoglein; multiple sclerosis;
KW	herpes simplex virus; adenovirus; phosphonamonometase;
KW	human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW	influenza; haemagglutinin; reovirus; sigma protein.
OS	Homo sapiens.
PN	W09637387-A1.
PD	12-SEP-1996.
DE	07-MAR-1996; U03182.

M P S E I (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch.p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:15:33 2000; Maspar time 401.42 Seconds

Tabular output not generated. 4.171 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150

Gap 15

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries
Maximum DB seq length 15

Database:

a-pending
1:PCT 2:06 3:060 4:07 5:080 6:081 7:082 8:083 9:084A
10:084B 11:085 12:086 13:087 14:088 15:089 16:090 17:091
18:092 19:093 20:094 21:095 22:NEWP 23:NEWU60 24:NEWU8
25:NEWU9

Statistics: Mean 21.138; Variance 58.541; scale 0.361

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	105	100.0	15	1	PCT-US96-0	Sequence 2, Applicatio
2	105	100.0	15	15	US-08-991-	Sequence 2, Applicatio

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	15 AA.
ID	PCT-US96-03182-2			
XX	xxxxxx			
AC				
XX				
DT				
XX				
DE	Sequence 2, Application PC/TUS9603182			
XX				
CC	Sequence 2, Application PC/TUS9603182			
CC	GENERAL INFORMATION:			

CC	APPLICANT:
CC	TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC	TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC	NUMBER OF SEQUENCES: 16
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC	STREET: 600 ATLANTIC AVENUE
CC	CITY: BOSTON
CC	STATE: MASSACHUSETTS
CC	COUNTRY: UNITED STATES OF AMERICA
CC	ZIP: 02210
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US96/03182
CC	FILING DATE:
CC	CLASSIFICATION:
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 08/400,796
CC	FILING DATE: 07-MAR-1995
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: GATES, EDWARD R.
CC	REGISTRATION NUMBER: 31,616
CC	REFERENCE/DOCKET NUMBER: H0498/7015MO
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 617-720-3500
CC	TELEFAX: 617-720-2441
CC	INFORMATION FOR SEQ ID NO: 2:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 15 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: peptide
CC	HYPOTHETICAL: NO
CC	FRAGMENT TYPE: Internal
CC	ORIGINAL SOURCE:
CC	ORGANISM: HOMO SAPIENS
CC	SEQUENCE 15 AA; 1652 MW; 1115 CN;
DB	1 FGIFVVDKNTGDINI 15
DB	1 FGIFVVDKNTGDINI 15
QY	1 FGIFVVDKNTGDINI 15
DE	Sequence 2, Application US/08991628
XX	
CC	Sequence 2, Application US/08991628
CC	GENERAL INFORMATION:
CC	APPLICANT: STROMINGER, Jack L
CC	APPLICANT: WUCHERPFENNIG, Kai
CC	TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC	TITLE OF INVENTION: IN AUTOIMMUNE DISEASE
CC	FILE REFERENCE: HAR-001DV
CC	CURRENT APPLICATION NUMBER: US/08/991,628
CC	CURRENT FILING DATE: 1997-11-15
CC	EARLIER APPLICATION NUMBER: USSN 08/400,796
CC	EARLIER FILING DATE: 1995-03-07
CC	NUMBER OF SEQ ID NOS: 20
CC	SOFTWARE: PatentIn Ver. 2.0

CC SEQ ID NO 2
 CC LENGTH: 15
 CC TYPE: PRT
 CC ORGANISM: Homo sapiens
 SQ SEQUENCE 15 AA: 1652 MW: 1115 CN:

Query Match 100.0%; Score 105; DB 15; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.57e-04;
 Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 FGIFVVDKNTGDINI 15
 QY 1 FGIFVVDKNTGDINI 15

Search completed: Sat May 13 05:22:31 2000
 Job time : 418 secs.

WORLDWIDE (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_DP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:14:06 2000; Maspar time 55.78 Seconds

Tabular output not generated. 3.484 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGTDINI 15

Scoring table: PAM 150

Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database: a-issued

1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 17.421; Variance 58.153; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	105	100.0	15	2	US-08-400- Sequence 2, Applicatio	2.01e-04

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-400-796-2 STANDARD; PRT: 15 AA.
XX
AC xxxxxx
XX
DT
XX
XX Sequence 2, Application US/08400796
XX
DE
CC Sequence 2, Application US/08400796
CC Patent No. 5874531
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, JACK L.
CC APPLICANT: WUCHERFENNIG, KAI
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF

CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/400,796
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: TWOMEY, MICHAEL J.
CC REGISTRATION NUMBER: 38,349
CC REFERENCE/DOCKET NUMBER: H0498/7015
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1652 MW: 1115 CN:
DB 1 FGIFVVDKNTGTDINI 15
Oy 1 FGIFVVDKNTGTDINI 15
Query Match 100.0%; Score 105; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.01e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: Sat May 13 05:15:11 2000
Job time : 65 secs.

 M P E S N E H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:40:34 2000; Maspar time 4.34 Seconds
 162.858 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
 Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: p1r62
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.968; Variance 38.706; scale 0.697.

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.

No matches found.

Search completed: Sat May 13 04:40:42 2000
 Job time : 8 secs.

WIDEORL (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:34:04 2000; Maspar time 101.68 Seconds

4.493 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table:
PAM 150
Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.553; Variance 34.618; scale 0.796

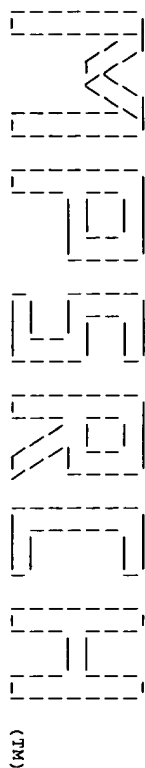
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	length	ID	Description
					Pred. No.

No matches found.

Search completed: Sat May 13 04:35:56 2000
Job time : 112 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:36:17 2000: Maspar time 223.46 seconds

Tabular output not generated. 4.654 Million cell updates/sec

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFFVVDKNTGDIINI 15

Scoring table:
PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.732: Variance 34.969: scale 0.764

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
.....						

No matches found.

Search completed: Sat May 13 04:40:15 2000
Job time : 238 secs.

THIS PAGE BLANK (USPTO)

Best Local Similarity 100.0%: Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
| | | | | | | | | | | | | | |
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 3

ID W78814 standard; peptide: 15 AA.
AC W78814;
DE 17-NOV-1998 (first entry)
DT Desmoglein 3 protein fragment 190-204.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
class II associated peptide; pathogen; gene therapy; genetic disease;
KM Infection; downregulation; immune response.
OS Homo sapiens.
OS Synthetic.
PN MO9831398-A1.
PD 23-JUL-1998.
PE 22-JAN-1998; US-003253.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR MPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
matrix and nucleic acid comprising an expression vector for use in
gene therapy
PT Disclosure: Page 8; 101pp; English.
PS A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (i) a
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 15 AA:

Query Match 100.0%; Score 96; DB 1; length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
| | | | | | | | | | | | | | |
Oy 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 04:48:26 2000
Job time : 8 secs.

 M P R S E R I E S

 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mprch_rp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 05:22:51 2000: Maspar time 52.56 Seconds
 3.698 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-3
 Description: (1-15) from US08991628.pep
 Perfect Score: 96
 Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table:
 PAM 150
 Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfile1

Statistics: Mean 16.952; Variance 50.735; scale 0.334

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description	Pred. No.
1	96	100.0	15	2 US-08-400- Sequence 3, Applicatio	2.73e-04
2	96	100.0	15	1 US-08-787- Sequence 24, Applicati	2.73e-04

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
 ID US-08-400-796-3 STANDARD: PRT: 15 AA.
 AC xxxxxx
 DT
 DE Sequence 3, Application US/08400796
 CC Sequence 3, Application US/08400796
 CC Patent No. 5874531
 CC GENERAL INFORMATION:
 CC APPLICANT: STROMINGER, JACK L.
 CC APPLICANT: WUCHERPFENNIG, KAI

CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
 CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
 CC NUMBER OF SEQUENCES: 16
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
 CC STREET: 600 ATLANTIC AVENUE
 CC CITY: BOSTON
 CC STATE: MA
 CC COUNTRY: USA
 CC ZIP: 02210
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/400,796
 CC FILING DATE:
 CC CLASSIFICATION: 424
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: TROMER, MICHAEL J.
 CC REGISTRATION NUMBER: 38,349
 CC REFERENCE/DOCKET NUMBER: H0498/7015
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617-720-3500
 CC TELEFAX: 617-720-2441
 CC INFORMATION FOR SEQ ID NO: 3:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL: NO
 CC FRAGMENT TYPE: Internal
 CC ORIGINAL SOURCE:
 CC ORGANISM: HOMO SAPIENS
 CC SEQUENCE 15 AA: 1645 MW; 1217 CN;
 CC
 CC Query Match 100.0%; Score 96; DB 2; Length 15;
 CC Best Local Similarity 100.0%; Pred. No. 2.73e-04;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 1 LNSKIAFKIVSOEPA 15
 CC 1 LNSKIAFKIVSOEPA 15
 CC
 CC RESULT 2
 ID US-08-787-547-24 STANDARD: PRT: 15 AA.
 AC xxxxxx
 DT
 DE Sequence 24, Application US/08787547
 CC Patent No. 5783567
 CC GENERAL INFORMATION:
 CC APPLICANT: Hedley, Mary Lynne
 CC APPLICANT: Curley, Joanne M.
 CC APPLICANT: Langer, Robert S.
 CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
 CC TITLE OF INVENTION: OF NOCLEIC ACID
 CC NUMBER OF SEQUENCES: 107
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Fish & Richardson, P. C.
 CC STREET: 225 Franklin Street
 CC CITY: Boston
 CC STATE: MA
 CC COUNTRY: US
 CC ZIP: 02110-2804
 CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM Compatible
 CC OPERATING SYSTEM: Windows95
 CC SOFTWARE: FastSeq for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/787,547
 CC FILING DATE: 22-JAN-1997
 CC CLASSIFICATION: 514
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER:
 CC FILING DATE:
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Fraser, Janis K.
 CC REGISTRATION NUMBER: 34,819
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617-542-5070
 CC TELEFAX: 617-542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ ID NO: 24:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 15 AA: 1645 MW: 1217 CN:

Query Match 100.00; Score 96; DB 1; Length 15;
 Best Local Similarity 100.00; Pred. No. 2.73e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 LNSKIAFKIVSOEPA 15
 QY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 05:23:57 2000
 Job time : 66 secs.

M P S R C H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:24:16 2000: Maspar time 398.53 seconds

Tabular output not generated. 4.202 Million cell updates/sec

Title: >US-08-991-628-3

Description: (1-15) from US08991628.pep

Perfect Score: 96

Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table: PAM 150

Gap 15

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database:

a-pending

1:PCT 2:U6 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A

10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91

18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEWU60 24:NEWU8

25:NEWU9

Statistics: Mean 20.540; Variance 50.756; scale 0.405

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	96	100.0	15	19	US-09-321- Sequence 24, Applicati	5.13e-04
2	96	100.0	15	16	US-09-003- Sequence 24, Applicati	5.13e-04
3	96	100.0	15	1	PCT-US96-0 Sequence 3, Applicatio	5.13e-04
4	96	100.0	15	15	US-08-991- Sequence 3, Applicatio	5.13e-04
5	96	100.0	15	1	PCT-US98-0 Sequence 24, Applicati	5.13e-04

Note: Post-processor removed 40 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT	ID	US-09-321-346-24	STANDARD:	PRT:	15 AA.
XX	XX	xxxxxx			
AC	XX				
XX	XX				
DT	XX				
DE	Sequence 24, Application US/09321346				

XX Sequence 24, Application US/09321346
CC GENERAL INFORMATION:
CC APPLICANT: Lunsford, Lynn B.
CC APPLICANT: Putnam, David
CC APPLICANT: Hedley, Mary Lynn
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC FILE REFERENCE: 08191/014001
CC CURRENT APPLICATION NUMBER: US/09/321,346
CC CURRENT FILING DATE: 1999-05-27
CC EARLIER APPLICATION NUMBER: US 09/266,463
CC EARLIER FILING DATE: 1999-03-11
CC NUMBER OF SEQ. ID NOS: 114
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 24
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SO
Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15
Query Match 100.0%; Score 96; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
ID US-09-003-253-24 STANDARD: PRT: 15 AA.
AC xxxxxx
DT
XX
DE Sequence 24, Application US/09003253
CC
CC GENERAL INFORMATION:
CC APPLICANT: Hedley, Mary Lynne
CC APPLICANT: Curley, Joanne M.
CC APPLICANT: Langer, Robert S.
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC NUMBER OF SEQUENCES: 108
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: Windows 95
CC SOFTWARE: FastSeq for Windows Version 2.0b
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/003,253
CC FILING DATE: 06-JAN-1998
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/035,983
CC FILING DATE: 22-JAN-1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Ph.D., J.D., Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 08191/003002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 24:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SO

Query Match 100.0%; Score 96; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 3
ID PCT-US96-03182-3 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 3, Application PC/TUS9603182
CC Sequence 3, Application PC/TUS9603182
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: H0498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: Internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SO

Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 4
ID US-08-991-628-3 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 3, Application US/08991628
CC Sequence 3, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L
CC APPLICANT: WUCHERPPENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC TITLE OF INVENTION: IN AUTOIMMUNE DISEASE
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 3
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
SO

Query Match 100.0%; Score 96; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.13e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
OY 1 LNSKIAFKIVSOEPA 15

RESULT 5
ID PCT-US98-01499-24 STANDARD; PRT: 15 AA.
XX xxxxxx
XX

Sequence 24, Application PC/TUS9801499
CC Sequence 24, Application PC/TUS9801499
CC GENERAL INFORMATION:
CC APPLICANT: Pangaea, Inc.
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY
CC TITLE OF INVENTION: OF NUCLEIC ACID
CC NUMBER OF SEQUENCES: 107
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson, P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: US
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: Windows95
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US98/01499
CC FILING DATE: 22-JAN-1998
CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/787,547
 CC FILING DATE: 22-JAN-1997
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Fraser, Janis K.
 CC REGISTRATION NUMBER: 34,819
 CC REFERENCE/DOCKET NUMBER: 08191/003WO1
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617-542-5070
 CC TELEFAX: 617-542-8906
 CC TELEX: 200154
 CC INFORMATION FOR SEQ ID NO: 24:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 15 AA: 1645 MW: 1217 CN:
 SO
 Query Match 100.0%; Score 96; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5,13e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 LNSKIAFKIVSOEPA 15
 |||||||||||||
 Oy 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 05:31:12 2000
 Job time : 416 secs.

WIDEORL (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_gp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 04:47:53 2000; MasPar time 4.31 Seconds
Tabular output not generated. 164.283 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pap
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.750; Variance 32.305; scale 0.797

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	No. Score Match	length DB	ID	Description	Pred. No.
No matches found.						

Search completed: Sat May 13 04:48:01 2000
Job time : 8 secs.

W E S E R C H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:41:28 2000; MasPar time 92.54 Seconds
4.936 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pap
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.327; Variance 28.416; scale 0.926

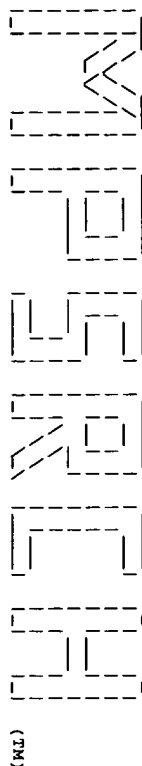
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
---------------	----------------	--------	----	-------------	-----------

No matches found.

Search completed: Sat May 13 04:43:12 2000
Job time : 104 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_gp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:43:31 2000; MasPar time 228.30 Seconds

Tabular output not generated. 4.555 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 225878 segs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: spiremb112
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.549; Variance 30.093; scale 0.849

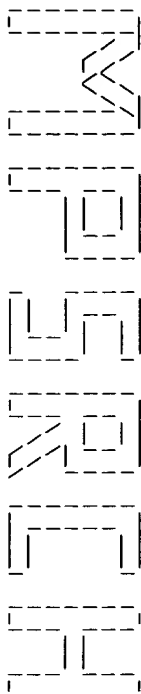
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description	Pred. No.
No matches found.					

Search completed: Sat May 13 04:47:34 2000
Job time : 243 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:33:05 2000; Maspar time 392.63 Seconds

Tabular output not generated. 4.265 Million cell updates/sec

Title: >US-08-991-628-4

Description: (1-15) from US08991628.pep

Sequence: 1 TPFLLSRNTGEVRT 15

Scoring table: PAM 150

Gap 15

Searched: 714217 segs, 111628056 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Maximum DB seq length 15

Database:

a-pending

1:PCR 2:U6 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A

10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91

18:U92 19:U93 20:U94 21:U95 22:NEWP 23:NEW60 24:NEW08

25:NEW09

Statistics: Mean 21.981; Variance 55.748; scale 0.394

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	100.0	15	15	US-08-991- Sequence 4, Applicatio	4.31e-05
2	108	100.0	15	1	PCT-US96-0 Sequence 4, Applicatio	4.31e-05
3	108	100.0	15	1	PCT-US98-0 Sequence 25, Applicati	4.31e-05
4	108	100.0	15	19	US-09-321- Sequence 25, Applicati	4.31e-05
5	108	100.0	15	16	US-09-003- Sequence 25, Applicati	4.31e-05
6	59	54.6	11	17	US-09-187- Sequence 3275, Applic	1.23e-02

Note: Post-processor removed 39 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT	ID	US-08-991-628-4	STANDARD:	PRT:	15 AA.
1	1	xxxxxx			
XX	XX				
DT	DT				
XX	XX				

DE	Sequence 4, Application US/08991628
XX	
CC	Sequence 4, Application US/08991628
CC	GENERAL INFORMATION:
CC	APPLICANT: STROMINGER, Jack L
CC	APPLICANT: WOCHERPFENNIG, Kai
CC	TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC	TITLE OF INVENTION: IN AUTOIMMUNE DISEASE
CC	FILE REFERENCE: HAR-001DV
CC	CURRENT APPLICATION NUMBER: US/08/991,628
CC	CURRENT FILING DATE: 1997-11-15
CC	EARLIER APPLICATION NUMBER: USSN 08/400,796
CC	EARLIER FILING DATE: 1995-03-07
CC	NUMBER OF SEQ ID NOS: 20
CC	SOFTWARE: PatentIn Ver. 2.0
CC	SEQ ID NO 4
CC	LENGTH: 15
CC	TYPE: PRT
CC	ORGANISM: Homo sapiens
SO	SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match 100.0%; Score 108; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1 TPFLLSRNTGEVRT 15
Oy	1 TPFLLSRNTGEVRT 15

RESULT	2
ID	PCT-US96-03182-4
XX	STANDARD:
XX	PRT: 15 AA.
XX	xxxxxx

DE	Sequence 4, Application PC/TUS9603182
CC	Sequence 4, Application PC/TUS9603182
CC	GENERAL INFORMATION:
CC	APPLICANT:
CC	TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC	TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC	NUMBER OF SEQUENCES: 16
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC	STREET: 600 ATLANTIC AVENUE
CC	CITY: BOSTON
CC	STATE: MASSACHUSETTS
CC	COUNTRY: UNITED STATES OF AMERICA
CC	ZIP: 02210
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US96/03182
CC	FILING DATE:
CC	CLASSIFICATION:
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 08/400,796
CC	FILING DATE: 07-MAR-1995
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: GATES, EDWARD R.
CC	REGISTRATION NUMBER: 31,616
CC	REFERENCE/DOCKET NUMBER: HQ498/7015WO
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 617-720-3500
CC	TELEFAX: 617-720-2441
CC	INFORMATION FOR SEQ ID NO: 4:
CC	SEQUENCE CHARACTERISTICS:

CC	LENGTH:	15 amino acids
CC	TYPE:	amino acid
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	peptide
CC	HYPOTHETICAL:	NO
CC	FRAGMENT TYPE:	Internal
CC	ORIGINAL SOURCE:	
CC	ORGANISM:	HOMO SAPIENS
S0	SEQUENCE	15 AA: 1722 MW; 1303 CN;
Db	Query Match	100.0%; Score 108; DB 1; Length 15;
	Best Local Similarity	100.0%; Pred. No. 4.31e-05;
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	1 TPMFLSRNTGEVRT 15	
	1 TPMFLSRNTGEVRT 15	
ID	RESULT 3	PRT: 15 AA.
XX	PCT-US98-01499-25	STANDARD;
AC	xxxxxx	
DT		
XX		
XX	Sequence 25, Application PC/TUS9801499	
DE	Sequence 25, Application PC/TUS9801499	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Pangaea, Inc.	
CC	TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY	
CC	TITLE OF INVENTION: OF NUCLEIC ACID	
CC	NUMBER OF SEQUENCES: 107	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Fish & Richardson, P.C.	
CC	STREET: 225 Franklin Street	
CC	CITY: Boston	
CC	STATE: MA	
CC	COUNTRY: US	
CC	ZIP: 02110-2804	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Diskette	
CC	OPERATING SYSTEM: Windows95	
CC	SOFTWARE: FASTSEQ for Windows Version 2.0	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: PCT/US98/01499	
CC	FILING DATE: 22-JAN-1998	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: 08/787,547	
CC	FILING DATE: 22-JAN-1997	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Fraser, Janis K.	
CC	REGISTRATION NUMBER: 34,819	
CC	REFERENCE/DOCKET NUMBER: 08191/003M01	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: 617-542-5070	
CC	TELEFAX: 617-542-8906	
CC	TELEX: 200154	
CC	INFORMATION FOR SEQ ID NO: 25:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 15 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: peptide	
S0	SEQUENCE	15 AA: 1722 MW; 1303 CN;
Query Match	100.0%; Score 108; DB 1; Length 15;	
Best Local Similarity	100.0%; Pred. No. 4.31e-05;	
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	. 1 TPMFLSRNTGEVRT 15	

```

0Y      1 TPFMFLSRNTEVRT 15
        |||
RESULT  4
ID US-09-321-346-25 STANDARD: PRT: 15 AA.
XX AC xxxxxx
XX DT
XX DE Sequence 25, Application US/09321346
CC CC Sequence 25, Application US/09321346
CC GENERAL INFORMATION:
CC APPLICANT: Lunsford, Lynn B.
CC APPLICANT: Putnam, David
CC APPLICANT: Hedley, Mary Lynn
CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC FILE REFERENCE: 08191/014001
CC CURRENT APPLICATION NUMBER: US/09/321,346
CC CURRENT FILING DATE: 1999-05-27
CC EARLIER APPLICATION NUMBER: US 09/266,463
CC EARLIER FILING DATE: 1999-03-11
CC NUMBER OF SEQ ID NOS: 14
CC SOFTWARE: FASTSEQ for Windows Version 3.0
CC SEQ ID NO: 25
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Homo sapiens
SQ SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match 100.0%; Score 108; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 4,31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps

Db 1 TPFMFLSRNTEVRT 15
   |||
OY 1 TPFMFLSRNTEVRT 15

RESULT  5
US-09-003-253 STANDARD: PRT: 15 AA.
XX AC xxxxxx
XX DT
XX DE Sequence 25, Application US/09003253
XX XX
DE DE Sequence 25, Application US/09003253
XX CC GENERAL INFORMATION:
CC CC APPLICANT: Hedley, Mary Lynne
CC CC APPLICANT: Curley, Joanne M.
CC CC APPLICANT: Langer, Robert S.
CC CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC CC NUMBER OF SEQUENCES: 108
CC CC CORRESPONDENCE ADDRESSES:
CC CC ADDRESSEE: Fish & Richardson P.C.
CC CC STREET: 225 Franklin Street
CC CC CITY: Boston
CC CC STATE: MA
CC CC COUNTRY: USA
CC CC ZIP: 02110-2804
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Diskette
CC CC COMPUTER: IBM Compatible
CC CC OPERATING SYSTEM: Windows 95
CC CC SOFTWARE: FASTSEQ for Windows Version 2.0b
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/09/003,253
CC CC FILING DATE: 06-JAN-1998
CC CC CLASSIFICATION: 514

```

```

CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 60/035,983
CC      FILING DATE: 22-JAN-1997
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Fraser, Ph.D., J.D., Janis K.
CC      REGISTRATION NUMBER: 34,819
CC      REFERENCE/DOCKET NUMBER: 08191/003002
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617/542-5070
CC      TELEFAX: 617/542-8906
CC      TELEX: 200154
CC      INFORMATION FOR SEQ ID NO: 25:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 15 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: 1linear
CC      MOLECULE TYPE: peptide
SO      SEQUENCE 15 AA; 1722 MW; 1303 CN;

Query Match      100.0%; Score 108; DB 16; Length 15;
Best Local Similarity 100.0%; Pred. No. 4,31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 TPFLLSRNTGEVRT 15
OY      1 TPFLLSRNTGEVRT 15

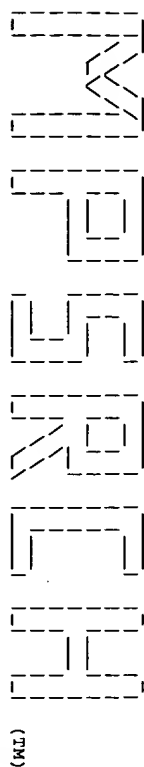
RESULT 6
ID      US-09-187-859-3275      STANDARD;      PRT;      11 AA.
XX      xxxxxx
XX
XX
DT
DT
XX
XX
DE      Sequence 3275, Application US/09187859
XX
CC      Sequence 3275, Application US/09187859
CC      GENERAL INFORMATION:
CC      APPLICANT: Blaschuk, Orest W.
CC      APPLICANT: Gour, Barbara J.
CC      TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
CC      TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
CC      FILE REFERENCE: 100086, 407C1
CC      CURRENT APPLICATION NUMBER: US/09/187,859
CC      CURRENT FILING DATE: 1998-11-06
CC      NUMBER OF SEQ ID NOS: 4052
CC      SOFTWARE: PatentIn Ver. 2.0
CC      SEQ ID NO 3275
CC      LENGTH: 11
CC      TYPE: PRT
CC      ORGANISM: Artificial Sequence
CC      FEATURE:
CC      OTHER INFORMATION: Representative cyclic modulating agent based on
CC      OTHER INFORMATION: Desmoglein cell adhesion recognition sequence
SO      SEQUENCE 11 AA; 1323 MW; 536 CN;

Query Match      54.6%; Score 59; DB 17; Length 11;
Best Local Similarity 70.0%; Pred. No. 1,23e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db      2 MFIINRNTGE 11
OY      3 MFIINRNTGE 12

```

Search completed: Sat May 13 05:39:53 2000
Job time : 408 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:53:44 2000; MasPar time 4.32 Seconds
163.716 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSRNTEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 27.219; Variance 35.537; scale 0.766

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	ID	Description
					Pred. No.

No matches found.

Search completed: Sat May 13 04:53:53 2000
Job time : 9 secs.

M P E R E I H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:48:45 2000; Mspar time 90.27 Seconds

Tabular output not generated. 5.061 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSRNTEGVRT 15

Scoring table: PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 27.727; Variance 32.559; scale 0.852

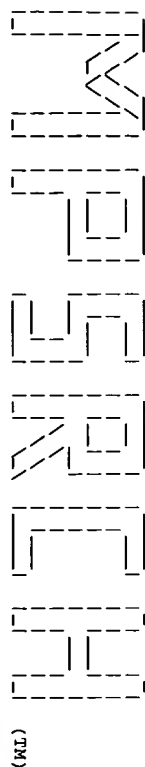
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	ID	Description
					Pred. No.

No matches found.

Search completed: Sat May 13 04:50:23 2000
Job time : 98 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:50:42 2000; Maspar time 149.09 Seconds
6.976 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPFLLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.384; Variance 31.843; scale 0.860

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.

No matches found.

Search completed: Sat May 13 04:53:24 2000
Job time : 162 secs.

THIS PAGE BLANK

 W04845

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 04:59:02 2000; Mpsrch time 3.04 Seconds
 Tabular output not generated. 116.885 Million cell updates/sec

Title: >US-08-991-628-5
 Description: (1-15) from US08991628.pap
 Perfect Score: 113
 Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table:
 PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: a-geneseg35
 1:genesegp

Statistics: Mean 18.468; Variance 59.453; scale 0.311

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	113	100.0	15	W04845	Self epitope of desmog 3.35e-05	
2	52	46.0	11	W13139	Human cadherin-5 antiq 1.72e+02	

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
 ID W04845 standard: peptide: 15 AA.
 AC W04845:
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.

PF 07-MAR-1996; U03182.
 PA 07-MAR-1995; US-400796.
 RA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Mucherpfennig KW;
 DR WPI: 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT autoimmune disease
 PS Claim 1; Page 40; 58pp; English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 251-285)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SQ Sequence 15 AA:






Query Match 100.0%; Score 113; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.35e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

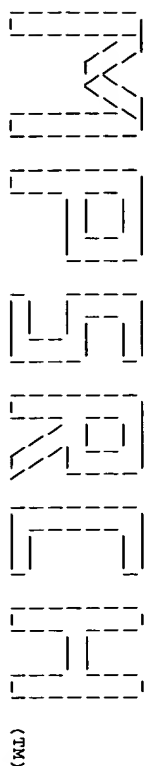
Db 1 CECNIKVKVDNDNFP 15
 |||||
 Oy 1 CECNIKVKVDNDNFP 15

RESULT 2
 ID W13139 standard: Protein; 11 AA.
 AC W13139:
 DT 14-MAY-1997 (first entry)
 DE Human cadherin-5 antigenic epitope (residues 242-252).
 KW Ca2+ dependent; cell adhesion protein; cadherin; human; antibody;
 KW purification; determination; epitope; tissue expression;
 KW binding antagonist; calcium ion; antigen.
 OS Homo sapiens.
 PN US5597725-A.
 PD 28-JAN-1997.
 PF 17-APR-1992; 872643.
 PR 17-APR-1992; US-872643.
 PR 19-APR-1993; US-049460.
 PR 26-JAN-1994; US-188228.
 PA (DOHE-) DOHENY EYE INSTR.
 PI Suzuki S;
 DR WPI: 97-108328/10.
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,
 PT etc.
 PS Claim 5; Column 112; 59pp; English.
 CC The present sequence is an antigenic epitope from human cadherin-5,
 CC which is a Ca2+ dependent cell adhesion protein. Antibodies or
 CC fragments that specifically bind the epitope can be used to purify
 CC the cadherin, determine its tissue expression and antagonise its
 CC ligand/antiligand binding activities.
 SQ Sequence 11 AA:

Query Match 46.0%; Score 52; DB 1; Length 11;
 Best Local Similarity 54.5%; Pred. No. 1.72e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Search completed: Sat May 13 04:59:09 2000
 Job time : 7 secs.



 Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 (TM)

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 05:41:35 2000; MasPar time 406.97 Seconds
 4.114 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-5
 Description: (1-15) from US08991628.pap
 Perfect Score: 113
 Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table:
 Gap 15
 PAM 150

Searched: 714217 seqs, 111628056 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database:
 a-pending
 1:PCT 2:06 3:U60 4:U7 5:U80 6:U81 7:U82 8:U83 9:U84A
 10:U84B 11:U85 12:U86 13:U87 14:U88 15:U89 16:U90 17:U91
 18:U92 19:U93 20:U94 21:U95 22:NEW06 23:NEW060 24:NEW08
 25:NEW09

Statistics: Mean 20.820; Variance 56.137; scale 0.371

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	113	100.0	15	1	PCT-US96-0	Sequence 5, Applicatio
2	113	100.0	15	15	US-08-991-	Sequence 5, Applicatio

Note: Post-processor removed 43 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
 ID PCT-US96-03182-5 STANDARD; PRT: 15 AA.
 AC xxxxxx
 DT
 DE Sequence 5, Application PC/TUS9603182
 CC Sequence 5, Application PC/TUS9603182
 CC GENERAL INFORMATION:

CC APPLICANT:
 CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
 CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
 CC NUMBER OF SEQUENCES: 16
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 CC STREET: 600 ATLANTIC AVENUE
 CC CITY: BOSTON
 CC STATE: MASSACHUSETTS
 CC COUNTRY: UNITED STATES OF AMERICA
 CC ZIP: 02210
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US96/03182
 CC FILING DATE:
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/400,796
 CC FILING DATE: 07-MAR-1995
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: GATES, EDWARD R.
 CC REGISTRATION NUMBER: 31,616
 CC REFERENCE/DOCKET NUMBER: H0498/7015WO
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 617-720-3500
 CC TELEFAX: 617-720-2441
 CC INFORMATION FOR SEQ ID NO: 5:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 15 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL: NO
 CC FRAGMENT TYPE: Internal
 CC ORIGINAL SOURCE:
 CC ORGANISM: HOMO SAPIENS
 CC SEQUENCE 15 AA: 1738 MW; 1181 CN;
 CC
 CC Query Match 100.0%; Score 113; DB 1; Length 15;
 CC Best Local Similarity 100.0%; Pred. No. 1.48e-05;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC Db 1 CECNIKVKVDNDNFP 15
 CC ||||||||||||
 CC Qy 1 CECNIKVKVDNDNFP 15
 CC
 CC RESULT 2
 CC ID US-08-991-628-5 STANDARD; PRT: 15 AA.
 CC AC xxxxxx
 CC DT
 CC DE Sequence 5, Application US/08991628
 CC CC GENERAL INFORMATION:
 CC APPLICANT: STROMINGER, Jack L
 CC APPLICANT: MUECHERFENNIG, Kai
 CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
 CC FILE REFERENCE: HAR-001DV
 CC CURRENT APPLICATION NUMBER: US/08/991,628
 CC CURRENT FILING DATE: 1997-11-15
 CC EARLIER APPLICATION NUMBER: USSN 08/400,796
 CC EARLIER FILING DATE: 1995-03-07
 CC NUMBER OF SEQ ID NOS: 20
 CC SOFTWARE: PatentIn Ver. 2.0



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:58:36 2000; Mspartime 4.28 Seconds
165.342 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNKKVQVNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.922; Variance 35.935; scale 0.721

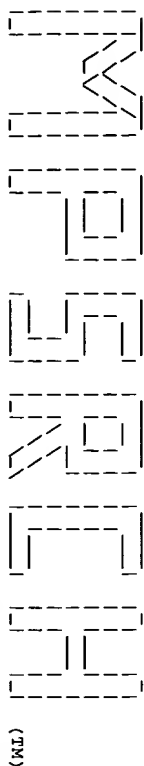
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
------------	-------------	--------	----	-------------	-----------

No matches found.

Search completed: Sat May 13 04:58:44 2000
Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:54:37 2000; MasPar time 79.85 Seconds
5.721 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNKKYKVDNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.676; Variance 31.455; scale 0.848

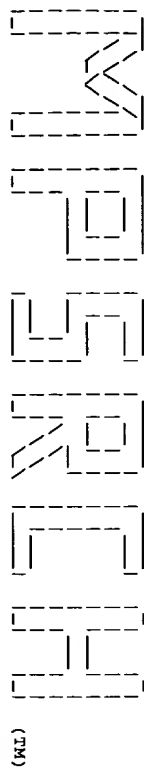
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description	Pred. No.
--------	-------	-------	--------	----	-------------	-----------

No matches found.

Search completed: Sat May 13 04:56:06 2000
Job time : 89 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mperch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:56:23 2000; MasPar time 105.85 Seconds
9.825 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIRKVDVNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.749; Variance 32.815; Scale 0.785

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

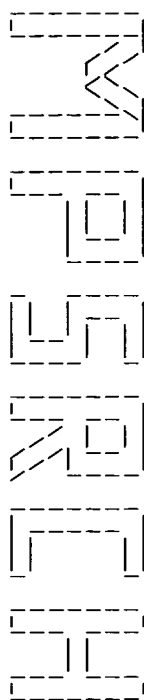
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
---------------	-------	----------------	--------	----	-------------	-----------

No matches found.

Search completed: Sat May 13 04:58:17 2000
Job time : 114 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:06:08 2000; Maspar time 3.03 Seconds
Tabular output not generated. 117.247 Million cell updates/sec

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pap
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-geneseg35
1:genesegp

Statistics: Mean 19.232; Variance 58.683; scale 0.328

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	109	100.0	15	1	W04846	Self epitope of desmog 5.72e-05	

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

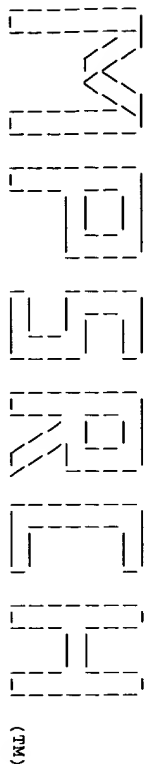
RESULT 1
ID W04846 standard; peptide: 15 AA.
AC W04846;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.

PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT autoimmune disease
PS Claim 1; Page 41; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 512-526)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 109; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 SARTLNRRYTGPTTF 15
QY 1 SARTLNRRYTGPTTF 15

Search completed: Sat May 13 05:06:15 2000
Job time : 7 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:49:00 2000; Maspar time 50.21 Seconds
3.871 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
(1-15) from US08991628.pep
Description: 108
Perfect Score: 1
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-1ssued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 18.236; Variance 59.295; scale 0.308

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

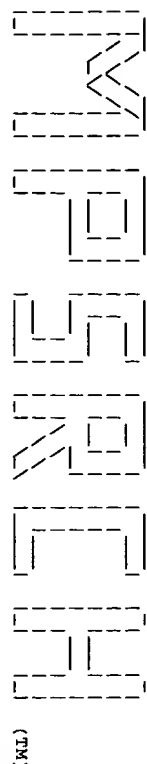
Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	109	100.0	15	2	US-08-400- Sequence 6, Applicatio	7.06e-05

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-400-796-6 STANDARD: PRT; 15 AA.
AC xxxxxx
XX
XX
DT
XX
XX
DE Sequence 6, Application US/08400796
CC
CC Patent No. 5874531
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, JACK L.
CC APPLICANT: WUCHERPFENNIG, KAI
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF

CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/400.796
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: TWOMEY, MICHAEL J
CC REGISTRATION NUMBER: 38,349
CC REFERENCE/DOCKET NUMBER: H0498/7015
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-3500
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA: 1761 MW: 1506 CN:
SQ
Query Match 100.0%; Score 109; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 7.06e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 SARTLNRRYTGPTTF 15
QY 1 SARTLNRRYTGPTTF 15
Search completed: Sat May 13 05:50:03 2000
Job time: 63 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:50:23 2000; Maspar time 402.79 seconds
4.157 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTF 15
Scoring table: PAM 150
Gap 15

Searched: 714217 seqs, 111628056 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-pending
1: PCT: 2: US 3: US60 4: U7 5: US80 6: US81 7: US82 8: US83 9: US84
10: US84B 11: US85 12: US86 13: US87 14: US88 15: US89 16: US90 17: US91
18: US92 19: US93 20: US94 21: US95 22: NEWP 23: NEWU60 24: NEWU8
25: NEWU9

Statistics: Mean 21.442; Variance 54.363; scale 0.394

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match Length	ID	Description
1	109	100.0	15 1	PCT-US96-0 Sequence 6, Applicatio
2	109	100.0	15 15	US-08-991- Sequence 6, Applicatio
3	56	51.4	10 19	US-09-361- Sequence 6, Applicatio

Note: Post-processor removed 42 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID PCT-US96-03182-6 STANDARD; PRT: 15 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 6, Application PC/TUS9603182
XX Sequence 6, Application PC/TUS9603182
CC

CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
CC TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
CC STREET: 600 ATLANTIC AVENUE
CC CITY: BOSTON
CC STATE: MASSACHUSETTS
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/03182
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/400,796
CC FILING DATE: 07-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: GATES, EDWARD R.
CC REGISTRATION NUMBER: 31,616
CC REFERENCE/DOCKET NUMBER: H0498/7015WO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-720-2441
CC TELEFAX: 617-720-2441
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: Internal
CC ORIGIN: SOURCE: Internal
CC ORGANISM: HOMO SAPIENS
CC SEQUENCE 15 AA; 1761 MW; 1506 CN;

Query Match 100.0%; Score 109; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.15e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SARTLNRRYTGPTF 15
|||||
1 SARTLNRRYTGPTF 15

RESULT 2
ID US-08-991-628-6 STANDARD; PRT: 15 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 6, Application US/08991628
XX
CC Sequence 6, Application US/08991628
CC GENERAL INFORMATION:
CC APPLICANT: STROMINGER, Jack L
CC APPLICANT: WUCHERPFENNIG, Kai
CC TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF ANTIGENS IMPLICATED
CC TITLE OF INVENTION: IN AUTOIMMUNE DISEASE
CC FILE REFERENCE: HAR-001DV
CC CURRENT APPLICATION NUMBER: US/08/991,628
CC CURRENT FILING DATE: 1997-11-15
CC EARLIER APPLICATION NUMBER: USSN 08/400,796
CC EARLIER FILING DATE: 1995-03-07
CC NUMBER OF SEQ ID NOS: 20

 MIPERIN
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 05:05:40 2000; MasPar time 4.33 Seconds
 Tabular output not generated. 163.439 Million cell updates/sec

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pep
 Perfect Score: 109
 Sequence: 1 SARTLNRYTGPTTF 15
 Scoring table: PAM 150
 Gap 15

Searched: 142080 segs, 47172406 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: p1r62
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 26.736; Variance 36.896; scale 0.725
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description	Pred. No.
No matches found.				

Search completed: Sat May 13 05:05:49 2000
 Job time : 9 secs.

 M P E R L H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 04:59:27 2000; Maspar time 83.07 Seconds
 5.499 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pep
 Perfect Score: 109
 Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
 PAM 150
 Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries
 Maximum DB seq length 15

Database: swiss-prot38
 1:swissprot

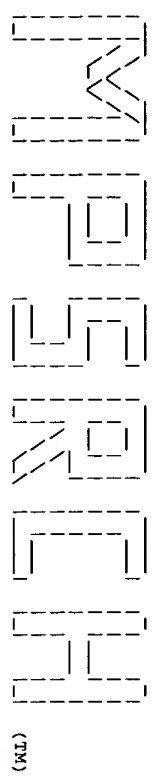
Statistics: Mean 27.244; Variance 33.691; scale 0.809

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description	Pred. No.
No matches found.				

Search completed: Sat May 13 05:01:04 2000
 Job time : 97 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp .protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:01:23 2000; Maspar time 224.41 Seconds
4.634 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARLNRRYTGPTTF 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_minc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_protent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.282; Variance 35.962; scale 0.731

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

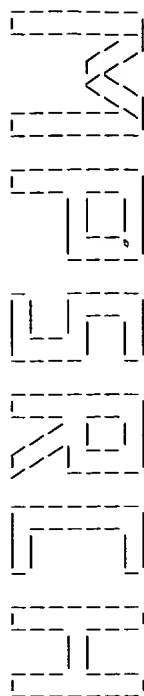
SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.

No matches found.

Search completed: Sat May 13 05:05:22 2000
Job time : 239 secs.

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:13:41 2000; Mspar time 3.01 Seconds
118.170 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) From US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table:
PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: a-geneseg35
1:genesegp

Statistics: Mean 17.183; Variance 54.742; scale 0.314

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description	Pred. No.
1	99	100.0	15	1 W04847 Self epitope of desmog 5.70e-04	

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID W04847 standard; peptide: 15 AA.
AC W04847;
DT 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerant; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.

PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Mucherpfennig KW;
DR WPI: 96-425218/42.
PR Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PR antigens - useful in disease treatment, and method for
PR identification of other self and non-self antigens implicated in
PR auto-immune disease.
PS Claim 1; Page 42; 56pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 763-786)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 99; DB 1; Length 15;
Best local similarity 100.0%; Pred. No. 5.70e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 QSGTMRTRHSTGCTN 15
QY 1 QSGTMRTRHSTGCTN 15

Search completed: Sat May 13 05:13:48 2000
Job time : 7 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 05:57:42 2000; Mapar time 47.18 seconds
Tabular output not generated. 4.120 Million cell updates/sec

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 OSGTMRTRHSTGCTN 15

Scoring table:
PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: a-1Issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfile1

Statistics: Mean 16.273; Variance 51.318; scale 0.317

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	99	100.0	15	2	US-08-400- Sequence 7, Applicatio	1.89e-04

Note: Post-processor removed 44 summaries from list due to search parameters chosen.

ALIGNMENTS

ID	Sequence	Standard	PRT	AA
US-08-400-796-7	STANDARD:			
xxxxxx				
Sequence 7, Application US/08400796				
Sequence 7, Application US/08400796				
Patent No. 5874531				
GENERAL INFORMATION:				
APPLICANT: STROMINGER, JACK L.				
APPLICANT: WOCHERFENNIG, KAI				
TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF				

TITLE OF INVENTION: ANTIGENS IMPLICATED IN AUTOIMMUNE DISEASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/400,796
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: TWOMEY, MICHAEL J.
REGISTRATION NUMBER: 38,349
REFERENCE/DOCKET NUMBER: H0498/7015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
SEQ SEQUENCE 15 AA; 1591 MW; 1253 CN;

Search completed: Sat May 13 05:58:40 2000
Job time : 58 secs.

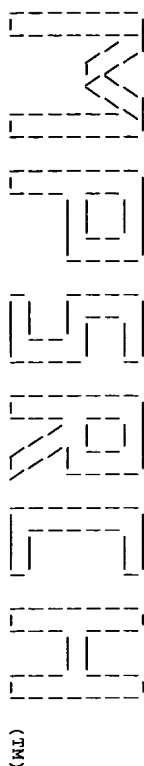
Query Match 100.0%; Score 99; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.89e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 OSGTMRTRHSTGCTN 15
QY 1 OSGTMRTRHSTGCTN 15

CC SEQ ID NO 7
 CC LENGTH: 15
 CC TYPE: PRT
 CC ORGANISM: Homo sapiens
 SQ SEQUENCE 15 AA: 1591 MW: 1253 CN:

Query Match 100.0%; Score 99; DB 15; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.97e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 OSGTMRTRHSTGCTN 15
 Oy 1 OSGTMRTRHSTGCTN 15

Search completed: Sat May 13 06:05:57 2000
 Job time : 415 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:13:15 2000; MasPar time 4.30 Seconds
164.392 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: pIr62
1:pIr1 2:pIr2 3:pIr3 4:pIr4

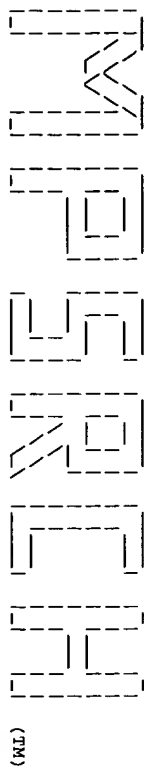
Statistics: Mean 23.943; Variance 29.203; scale 0.820

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description	Pred. No.
No matches found.					

Search completed: Sat May 13 05:13:24 2000
Job time : 9 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 05:06:34 2000; MasPar time 91.30 Seconds
5.004 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 83857 segs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries
Maximum DB seq length 15

Database: swiss-prot38
1:swissprot

Statistics: Mean 24.519; Variance 27.183; scale 0.902

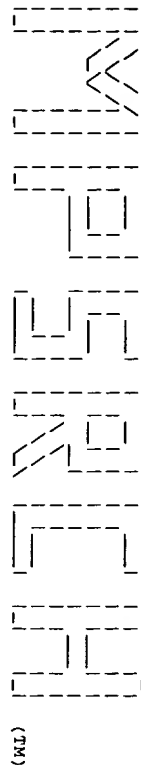
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length DB	ID	Description	Pred. No.

No matches found.

Search completed: Sat May 13 05:08:16 2000
Job time : 102 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Sat May 13 05:08:36 2000; MasPar time 245.94 Seconds
4.229 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRTRHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries
Maximum DB seq length 15

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.224; Variance 25.706; scale 0.942

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
No matches found.						

Search completed: Sat May 13 05:12:56 2000
Job time : 260 secs.

 W04842
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:05:15 2000: Maspar time 3.02 Seconds
 Tabular output not generated. 117.657 Million cell updates/sec

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 18.538; Variance 60.235; scale 0.308

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	105	100.0	15	W04842	self epitope of desmog	3.51e-04
2	105	100.0	30	R93960	peptide combining with	3.51e-04
3	105	100.0	614	W07908	pemphigus vulgaris ant	3.51e-04
4	105	100.0	999	R30742	Human pemphigus vulgar	3.51e-04
5	86	81.9	19	R93961	peptide combining with	4.92e-02
6	86	81.9	21	R93962	peptide combining with	4.92e-02
7	84	80.0	778	W15489	pemphigus foliaceus an	8.20e-02
8	82	78.1	778	W15489	Segment of desmosomal	1.36e-01
9	57	54.3	187	Y11097	H. pylori ORF hp7e1043	5.92e-01
10	57	54.3	204	Y11098	H. pylori ORF hp7e1043	5.92e-01
11	57	54.3	204	W98278	H. pylori GHP 876 pro	5.92e-01
12	57	54.3	693	R49731	Sequence encoded by hu	5.92e-01
13	57	54.3	796	R49732	Sequence encoded by hu	5.92e-01
14	57	54.3	796	W25636	Human cadherin-11.	5.92e-01
15	57	54.3	796	R49730	Sequence encoded by mu	5.92e-01
16	57	54.3	796	W85598	Cadherin-11.	5.92e-01
17	57	54.3	796	W13134	Full length human can	5.92e-01
18	57	54.3	797	R86866	Human protocadherin pc	5.92e-01
19	57	54.3	1400	W82791	Human RON receptor pro	7.45e-01
20	56	53.3	433	W41732	Arabidopsis chloroplast	7.45e-01
21	56	53.3	511	R69506	Aspergillus sp. recomb	7.45e-01
22	56	53.3	515	R69505	Aspergillus sp. recomb	7.45e-01
23	56	53.3	515	R69504	Aspergillus sp. recomb	7.45e-01

24	56	53.3	713	1	W25638	Human cadherin-13.	7.45e-01
25	56	53.3	713	1	W13136	Full length human cadh	7.45e-01
26	55	52.4	17	1	R93963	Peptide combining with	9.36e-01
27	55	52.4	20	1	W4250	Human wild-type E-cadh	9.36e-01
28	55	52.4	499	1	R60609	Tobamovirus replicatio	9.36e-01
29	55	52.4	878	1	R55060	Sequence of human live	9.36e-01
30	55	52.4	878	1	R85487	Human E-cadherin precu	9.36e-01
31	55	52.4	1827	1	W74090	Human hsi protein sequ	9.36e-01
32	54	51.4	246	1	R25789	Sequence encoded by th	1.18e-02
33	54	51.4	246	1	W76815	N. gonorrhoeae pilQ2 pr	1.18e-02
34	54	51.4	557	1	W43448	Tobacco laccase clone	1.18e-02
35	53	50.5	199	1	R34398	Helicobacter pylori ur	1.47e-02
36	53	50.5	794	1	W25637	Human cadherin-12.	1.47e-02
37	53	50.5	794	1	W13135	Putative human cadheri	1.47e-02
38	53	50.5	2647	1	W13349	Human filamin.	1.47e-02
39	52	49.5	115	1	W62680	Streptococcus pneumoni	1.84e-02
40	52	49.5	507	1	W29772	Malassezia fungus MF-5	1.84e-02
41	52	49.5	535	1	R56549	Cold acclimatization p	1.84e-02
42	52	49.5	1026	1	R58906	Human protocadherin-42	1.84e-02
43	52	49.5	1026	1	R87146	Protocadherin clone 42	1.84e-02
44	52	49.5	1203	1	R58911	Product of alternative	1.84e-02
45	52	49.5	1203	1	R87152	Alternatively spliced	1.84e-02

ALIGNMENTS

RESULT 1
 ID W04842 standard; peptide: 15 AA.
 AC W04842:
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphononmutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W0962787-AL.
 PD 12-SEP-1996.
 PR 07-MAR-1996; U03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger Jr., Muecherfennig KW:
 DR WPT: 96-425218/42.
 PT pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 38: 58pp: English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 97-111)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SO Sequence 15 AA:

Query Match 100.0%; Score 105; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.51e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FGIFVVDKNTGDINI 15
 QY 1 FGIFVVDKNTGDINI 15

RESULT 2

ID R93960 standard; peptide: 30 AA.

AC R93960: (first entry)
 DT 05-JUL-1996 (first entry)
 DE Peptide combining with anti-Interepidermal cellular antibody.
 KW anti-Interepidermal cellular antibody; autoantibody; adsorbent.
 OS Synthetic.
 PN J07309893-A.
 PD 28-NOV-1995.
 PF 18-MAY-1994: 129556.
 PR 18-MAY-1994: JP-129556.
 PA (KURS) KURARAY CO LTD.
 DR WPI: 96-045392/05.
 PT Anti-Interepidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody.
 PS Claim 1: Page 2: 7pp; Japanese.
 CC A new peptide is disclosed which contains at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asp Thr Ala Ile Val Asp Arg Glu Glu (the present sequence), the peptide not containing more than 50 residues. The peptide combines with anti-Interepidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-Interepidermal cellular antibody.
 SO Sequence 30 AA;

Query Match 100.0%; Score 105; DB 1; Length 30;

Best Local Similarity 100.0%; Pred. No. 3.51e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

DB 8 FGIFVVDKNTGDI 22

QY 1 FGIFVVDKNTGDI 15

RESULT 3

ID W07908 standard; protein: 614 AA.
 AC W07908:
 DT 29-JAN-1997 (first entry)
 DE Pemphigus vulgaris antigen protein extracellular region.
 KW Autoantibody; immunoglobulin G; IgG; fusion protein; diagnosis; treatment; pemphigus vulgaris; PV; bulla; blister; skin disease; dermatology.
 OS Homo sapiens.
 PN J08188540-A.
 PD 23-JUL-1996.
 PF 30-JUN-1995: 165632.
 PR 30-JUN-1994: JP-173291.
 PA (NISH) NISHUKAWA T.
 DR WPI: 96-388562/39.
 PT Fused protein recognised by pemphigus vulgaris autoantibody - useful to treat and diagnose pemphigus vulgaris
 PS Claim 1: Page 7-9: 9pp; Japanese.
 CC W07908 represents the human pemphigus vulgaris (PV) antigen extracellular region. The PV antigen is produced in patients with pemphigus vulgaris resulting in autoimmune disease. PV is a rare relapsing disease causing suprabasal, intra-epidermal bullae (vesicles) of the skin and mucous membranes, which is fatal if untreated. The PV antigen was fused to a human IgG1 hinge region and the resulting fusion protein is useful to treat or diagnose pemphigus vulgaris.
 SO Sequence 614 AA;

Query Match 100.0%; Score 105; DB 1; Length 614;

Best Local Similarity 100.0%; Pred. No. 3.51e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

DB 96 FGIFVVDKNTGDI 110

QY 1 FGIFVVDKNTGDI 15

RESULT 4

ID R30742 standard; protein: 999 AA.

AC R30742:
 DT 14-JUN-1993 (first entry)
 DE Human pemphigus vulgaris 130KD antigen.
 KW Pemphigus vulgaris; skin disease; autoantibodies; keratinocyte cell surface antigen; glycoprotein; cell adhesion.
 OS Homo sapiens.
 PN US7798918-A.
 PD 15-DEC-1992.
 PF 27-NOV-1991: 798918.
 PR 27-NOV-1991: US-798918.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 DR Amagai M, Klaus-Kovtun V, Stanley JR;
 DR WPI: 93-067436/08.
 DR N-PDB: Q35992.
 PT DNA encoding pemphigus vulgaris antigen - useful in proteins for diagnostic and therapeutic uses.
 PS Disclosure: Fig 7: 50pp; English.
 CC This sequence is the pemphigus vulgaris 130KD antigen. The protein and its encoding DNA may be used in the diagnosis and treatment of pemphigus vulgaris. It is thought that the antigen may be a cell adhesion molecule.
 SO Sequence 999 AA;

Query Match 100.0%; Score 105; DB 1; Length 999;

Best Local Similarity 100.0%; Pred. No. 3.51e-04; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

DB 97 FGIFVVDKNTGDI 111

QY 1 FGIFVVDKNTGDI 15

RESULT 5

ID R93961 standard; peptide: 19 AA.
 AC R93961:
 DT 05-JUL-1996 (first entry)
 DE Peptide combining with anti-Interepidermal cellular antibody.
 KW anti-Interepidermal cellular antibody; autoantibody; adsorbent.
 OS Synthetic.
 PN J07309893-A.
 PD 28-NOV-1995.
 PF 18-MAY-1994: 129556.
 PR 18-MAY-1994: JP-129556.
 PA (KURS) KURARAY CO LTD.
 DR WPI: 96-045392/05.
 PT Anti-Interepidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody.
 PS Example 1: Page 4: 7pp; Japanese.
 CC New peptides are disclosed which contain at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asp Thr Ala Ile Val Asp Arg Glu Glu (see R93960), the peptide not containing more than 50 residues. The peptide combines with anti-Interepidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-Interepidermal cellular antibody.
 SO Sequence 19 AA;

Query Match 81.9%; Score 86; DB 1; Length 19;

Best Local Similarity 100.0%; Pred. No. 4.92e-02; Mismatches 0; Indels 0; Gaps 0;

Matches 12; Conservative 0;

DB 8 FGIFVVDKNTGD 19

QY 1 FGIFVVDKNTGD 12

RESULT 6

ID R93962 standard; peptide: 21 AA.

AC R93962;
 DT 05-JUL-1995 (first entry)
 DE Peptide combining with anti-interepidermal cellular antibody.
 KW anti-interepidermal cellular antibody; autoantibody; adsorbent.
 OS Synthetic.
 PN J07309893-A.
 PD 28-NOV-1995.
 PF 18-MAY-1994: 129556.
 PR 18-MAY-1994: JP-129556.
 PA (KURS.) KURARAY CO LTD.
 WP: 96-045392/05.
 DT Anti-interepidermal cellular antibody-combining peptide - which can be immobilised on a column to form adsorbent useful for treating diseases related to the antibody.
 PT Example 2: Page 5: 7pp: Japanese.
 PS New peptides are disclosed which contain at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg Glu Glu (see R93960), the peptide not containing more than 50 residues. The peptide combines with anti-interepidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-interepidermal cellular antibody. The present sequence is a specific example of the new peptides.
 CC Sequence 21 AA;

Query Match 81.9%; Score 86; DB 1; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.92e-02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 FGIFVNDKNTGD 21
 OY 1 FGIFVNDKNTGD 12

RESULT 7
 ID W15489 standard; Protein: 778 AA.
 AC W15489;
 DT 17-JUN-1997 (first entry)
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;
 KW extracellular region; antigen; hinge portion; skin;
 KW dermatitis herpetiformis; fusion protein; detection; ss.
 OS Chimeric - Homo sapiens.
 FH Key Location/Qualifiers
 FT domain 1..545
 PN J09077800-A.
 PD 25-MAR-1997.
 PF 12-SEP-1995: 260899.
 PR 12-SEP-1995: JP-260899.
 PA (NISHU.) NISHIKAWA T.
 DR WP: 97-241758/22.
 P-PSDB: T66428.
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked through the hinge region used to treat pemphigus foliaceus
 PS Claim 1: Page 10-12: 17pp: Japanese.
 CC This sequence represents a fused protein recognised by pemphigus foliaceus patient autoantibody which comprises the constant region of IgG linked to the extracellular region of pemphigus foliaceus antigen protein through the hinge portion. Pemphigus foliaceus is a chronic, generalised, vesicular and scaling skin eruption similar to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion protein is useful to treat pemphigus foliaceus. The antigen is especially administered through an adsorbent upon which the fusion protein is immobilised via a carrier. The fusion protein is also useful for detecting pemphigus foliaceus antibodies which is useful in immunodiagnosis. The fusion protein has little or no side effects.
 CC Sequence 778 AA;

Query Match 80.0%; Score 84; DB 1; Length 778;
 Best Local Similarity 60.0%; Pred. No. 8.20e-02;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVNDKTEINI 111
 OY 1 FGIFVNDKNTGDINI 15

RESULT 8
 ID W13009 standard; Protein: 560 AA.
 AC W13009;
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2.
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
 KW micrometastasis; separation; enrichment; targeted delivery;
 KW metastatic.
 OS Homo sapiens.
 PN DE19531033-A1.
 PD 27-FEB-1997.
 PF 23-AUG-1995: 031033.
 PR 23-AUG-1995: DE-031033.
 PA (PROG-) PROGEN BIOTECHNIK GMBH.
 PI Franke WW, Schaefer S:
 WP: 97-146518/14.
 DT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases
 PS Claim 7: Page 5: 8pp: German.
 CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinoma.
 CC Sequence 560 AA;

Query Match 78.1%; Score 82; DB 1; Length 560;
 Best Local Similarity 60.0%; Pred. No. 1.36e-01;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 50 FGIFVNDKTEINI 64
 OY 1 FGIFVNDKNTGDINI 15

RESULT 9
 ID Y11097 standard; Protein: 187 AA.
 AC Y11097;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF hp/pe10433_36339535_f3.3 secreted protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein.
 KW secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN M09824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997: U22104.
 PR 14-JUL-1997: US-891928.
 PR 05-DEC-1996: US-759625.
 PR 25-MAR-1997: US-823745.
 PA (ASPR.) ASTRA AB.
 PI Alm RA, Castriotta LM, Dolg PC, Kabok Z, Smith D:
 WP: 96-33501/29.
 P-PSDB: X30626.
 DT New isolated Helicobacter pylori nucleic acids - used to develop products for the diagnosis, prevention and treatment of infection by H. pylori and other Helicobacter species
 PS Claims 37, 41: Page 264-265: 339pp: English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides are disclosed, together with the nucleic acids encoding them. In all,

CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 187 AA.

Query Match 54.3%; Score 57; DB 1; Length 187;
 Best Local Similarity 46.2%; Pred. No. 5,92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 67 GFFVVKKSANEID 79
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 10
 ID Y11098 standard; Protein: 204 AA.

AC Y11098:1999 (first entry)
 DE H. pylori ORF hp7el0433_36339535_63_3 secreted protein.
 KW Vaccine: probe; diagnostic; ORF; cell envelope protein.
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN M09824475-A1.
 PD 11-JUN-1998.
 PE 05-DEC-1997; U22104.
 PR 14-JUL-1997; US-891928.
 PR 05-DEC-1996; US-759625.
 PR 25-MAR-1997; US-823745.
 PA (ASPR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 DR WPI: 98-333051/29.
 DR N-PSDB: X10627.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41: Page 265-266: 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 204 AA.

Query Match 54.3%; Score 57; DB 1; Length 204;
 Best Local Similarity 46.2%; Pred. No. 5,92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 62 GFFVVKKSANEID 74
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 11
 ID W98278 standard; Protein: 204 AA.

AC W98278:
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 876 protein.
 KM GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 OS Helicobacter pylori.
 PN M09843478-A1.

PD 08-OCT-1998.
 PE 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 DR Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI: 98-542293/46.
 DR N-PSDB: X13997.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8: Page 315-316: 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHPD protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SO Sequence 204 AA.

Query Match 54.3%; Score 57; DB 1; Length 204;
 Best Local Similarity 46.2%; Pred. No. 5,92e+01;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 62 GFFVVKKSANEID 74
 Oy 2 GFFVVDKNTGIDIN 14

RESULT 12
 ID R49732 standard; Protein: 693 AA.
 AC R49732:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-2 CDNA.
 KW OSF-4-2; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KM diagnosis.
 OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PE 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 PA (EARH) HOECHST JAPAN LTD.
 PI Amano E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: Q44393.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 PT and diagnosis of bone metabolic disease, and nucleic acid
 PT encoding them
 PS Claim 1: Page 23-27: 34pp; English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 CC then as much common DNA as possible removed by hybridisation between
 CC the 2 libraries. Residual E1-specific DNA was amplified, inserted
 CC into lambda g10 and screened by plaque hybridisation. A minibank of
 CC 273 E-specific clones was recovered, their inserts amplified and
 CC used to screen total RNA from both cell types. One clone specific
 CC for E1 was identified and sequenced. The insert from this clone was
 CC used to screen cDNA prep. from E1 RNA and the longest posn. insert
 CC cloned in pGEM 112f (+) to give pKOR164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SO Sequence 693 AA;

Query Match 54.3%; Score 57; DB 1; Length 693;
 Best Local Similarity 50.0%; Pred. No. 5,92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGNTH 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 13
 ID R49731 standard; Protein: 796 AA.
 AC R49731:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-1 cDNA.
 KW OSF-4-1; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KW diagnosis.
 OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 PA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR N-PSDB: 044392.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 PT and diagnosis of bone metabolic disease, and nucleic acid
 PT encoding them
 PS Claim 1: Page 18-22: 34pp: English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 CC then as much common DNA as possible removed by hybridisation between
 CC the 2 libraries. Residual EI-specific DNA was amplified, inserted
 CC into lambda gt10 and screened by plaque hybridisation. A minbank of
 CC 273 E-specific clones was recovered, their inserts amplified and
 CC used to screen total RNA from both cell types. One clone specific
 CC for EI was identified and sequenced. The insert from this clone was
 CC used to screen cDNA prep. from EI RNA and the longest posn. insert
 CC cloned in pGEM 112f (+) to give PKO164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGNTH 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 14
 ID W25636 standard; Protein: 796 AA.
 AC W25636:
 DT 03-NOV-1997 (first entry)
 DE Human cadherin-11.
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;
 KW superfamily; cytoskeleton; eatenin; cancer.
 OS Homo sapiens.
 PN US5646250-A.
 PD 08-JUL-1997.
 PF 17-APR-1992; 872643.
 PR 19-APR-1993; US-049460.
 PR 17-APR-1992; US-872643.
 PR 01-NOV-1994; US-332638.
 PA (DOHE-) DOHEMY EYE INST.
 PI Suzuki S;
 DR WPI: 97-362997/33.
 DR N-PSDB: T85403.
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
 PS Claim 1: Column 89-94: 56pp: English.
 CC This sequence represents human cadherin-11. The invention specifically
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat

CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell
 CC adhesion proteins. They are glycosylated integral membrane proteins
 CC that have an N-terminal extracellular domain that determines binding
 CC specificity, a hydrophobic membrane spanning region and a C-terminal
 CC cytoplasmic domain, which is highly conserved among members of the
 CC superfamily. The C-terminal domain interacts with the cytoskeleton
 CC through eatens and other cytoskeleton-associated proteins. The
 CC novel cadherin proteins may be used in the analysis of the role of
 CC cadherins in various cancers. Sequence analysis of the cadherin
 CC proteins also allows investigation of the structure and function of
 CC cadherin. The cadherin proteins may be isolated by using anti-cadherin
 CC antibodies. These antibodies may also be used to modulate the activity
 CC of cadherin and to determine the tissue specific distribution of cadherin
 CC proteins. Each subclass of cadherins has a unique tissue distribution
 CC pattern.
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGNTH 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 15
 ID R49730 standard; Protein: 796 AA.
 AC R49730:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by murine OSF-4 cDNA.
 KW OSF-4; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KW diagnosis.
 OS Mus musculus.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 PA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: Q44391.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 PT and diagnosis of bone metabolic disease, and nucleic acid
 PT encoding them
 PS Claim 1: Page 13-17: 34pp: English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 CC then as much common DNA as possible removed by hybridisation between
 CC the 2 libraries. Residual EI-specific DNA was amplified, inserted
 CC into lambda gt10 and screened by plaque hybridisation. A minbank of
 CC 273 E-specific clones was recovered, their inserts amplified and
 CC used to screen total RNA from both cell types. One clone specific
 CC for EI was identified and sequenced. The insert from this clone was
 CC used to screen cDNA prep. from EI RNA and the longest posn. insert
 CC cloned in pGEM 112f (+) to give PKO164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prep. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SO Sequence 796 AA:

Query Match 54.3%; Score 57; DB 1; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 100 IFVIDKSGNTH 111
 |||:|::|:|:
 Oy 3 IFVVDKNTGDIN 14

RESULT 16
 ID W25636 standard; Protein: 796 AA.
 AC W25636:
 DT 03-NOV-1997 (first entry)
 DE Human cadherin-11.
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;
 KW superfamily; cytoskeleton; eatenin; cancer.
 OS Homo sapiens.
 PN US5646250-A.
 PD 08-JUL-1997.
 PF 17-APR-1992; 872643.
 PR 19-APR-1993; US-049460.
 PR 17-APR-1992; US-872643.
 PR 01-NOV-1994; US-332638.
 PA (DOHE-) DOHEMY EYE INST.
 PI Suzuki S;
 DR WPI: 97-362997/33.
 DR N-PSDB: T85403.
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
 PS Claim 1: Column 89-94: 56pp: English.
 CC This sequence represents human cadherin-11. The invention specifically
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat

Search completed: Sat May 13 07:05:22 2000

Mon May 15 08:08:02 2000

US-08-991-628-2a.rag

Page 5

Job time : 7 secs.

1. *Pharmaceuticals*
 2. *Medical Devices*
 3. *Biotechnology*
 4. *Healthcare Services*
 5. *Medical Research*
 6. *Health Insurance*
 7. *Medical Education*
 8. *Healthcare Policy*
 9. *Medical Ethics*
 10. *Healthcare Economics*
 11. *Medical Law*
 12. *Healthcare Management*
 13. *Medical History*
 14. *Healthcare Innovation*
 15. *Medical Practice*
 16. *Healthcare Regulation*
 17. *Medical Research Ethics*
 18. *Healthcare Quality Improvement*
 19. *Medical Device Regulation*
 20. *Healthcare Access*
 21. *Medical Device Innovation*
 22. *Healthcare Equity*
 23. *Medical Device Safety*
 24. *Healthcare Sustainability*
 25. *Medical Device Design*
 26. *Healthcare Reform*
 27. *Medical Device Testing*
 28. *Healthcare Financing*
 29. *Medical Device Approval*
 30. *Healthcare Delivery*
 31. *Medical Device Evaluation*
 32. *Healthcare Policy Analysis*
 33. *Medical Device Research*
 34. *Healthcare Law*
 35. *Medical Device Development*
 36. *Healthcare Ethics*
 37. *Medical Device Innovation*
 38. *Healthcare Economics*
 39. *Medical Device Regulation*
 40. *Healthcare Management*
 41. *Medical Device History*
 42. *Healthcare Innovation*
 43. *Medical Device Practice*
 44. *Healthcare Regulation*
 45. *Medical Device Research Ethics*
 46. *Healthcare Quality Improvement*
 47. *Medical Device Regulation*
 48. *Healthcare Access*
 49. *Medical Device Innovation*
 50. *Healthcare Equity*
 51. *Medical Device Safety*
 52. *Healthcare Sustainability*
 53. *Medical Device Design*
 54. *Healthcare Reform*
 55. *Medical Device Testing*
 56. *Healthcare Financing*
 57. *Medical Device Approval*
 58. *Healthcare Delivery*
 59. *Medical Device Evaluation*
 60. *Healthcare Policy Analysis*
 61. *Medical Device Research*
 62. *Healthcare Law*
 63. *Medical Device Development*
 64. *Healthcare Ethics*
 65. *Medical Device Innovation*
 66. *Healthcare Economics*
 67. *Medical Device Regulation*
 68. *Healthcare Management*
 69. *Medical Device History*
 70. *Healthcare Innovation*
 71. *Medical Device Practice*
 72. *Healthcare Regulation*
 73. *Medical Device Research Ethics*
 74. *Healthcare Quality Improvement*
 75. *Medical Device Regulation*
 76. *Healthcare Access*
 77. *Medical Device Innovation*
 78. *Healthcare Equity*
 79. *Medical Device Safety*
 80. *Healthcare Sustainability*
 81. *Medical Device Design*
 82. *Healthcare Reform*
 83. *Medical Device Testing*
 84. *Healthcare Financing*
 85. *Medical Device Approval*
 86. *Healthcare Delivery*
 87. *Medical Device Evaluation*
 88. *Healthcare Policy Analysis*
 89. *Medical Device Research*
 90. *Healthcare Law*
 91. *Medical Device Development*
 92. *Healthcare Ethics*
 93. *Medical Device Innovation*
 94. *Healthcare Economics*
 95. *Medical Device Regulation*
 96. *Healthcare Management*
 97. *Medical Device History*
 98. *Healthcare Innovation*
 99. *Medical Device Practice*
 100. *Healthcare Regulation*

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MSEARCH protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:04:49 2000; Maspar time 4.29 Seconds
Tabular output not generated.
164.911 Million cell updates/sec
(TW)

Title: >US-08-991-628-2
Description: (1-15) from US08991628.pep
Perfect Score: 105
Sequence: 1 FGIFVVDKNTGDINI 15
Scoring table: PAM 150
Gap 15
Searched: 142080 seqs, 47172406 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4
Statistics: Mean 26.968; Variance 38.706; scale 0.697

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	105	100.0	999	1	IUHUG3	desmoglein 3 precursor	1.33e-09
2	84	80.0	1043	1	IUBOG1	desmoglein 1 precursor	6.14e-05
3	84	80.0	1049	1	IUHUG1	desmoglein 1 precursor	6.14e-05
4	82	78.1	1117	2	S38673	desmoglein 2 - human	1.63e-04
5	66	62.9	761	1	IUBODE	desmoglein 1a - bovl	2.64e-01
6	66	62.9	785	2	I50180	cadherin-7 - chicken	2.64e-01
7	66	62.9	839	1	IJBODF	desmocollin 1b precursor	2.64e-01
8	65	61.9	770	2	I48910	desmocollin 1b precursor	4.08e-01
9	65	61.9	824	2	A48910	desmocollin 1a precursor	4.08e-01
10	65	61.9	840	2	I37281	Desclb precursor - hum	4.08e-01
11	65	61.9	894	2	I37282	Desclb precursor - hum	4.08e-01
12	61	58.1	1612	2	S59169	DNA topoisomerase (AT	2.22e+00
13	61	58.1	1626	2	A39242	DNA topoisomerase (AT	2.22e+00
14	60	57.1	94	2	A65141	hypothetical 10.6 kD	3.36e+00
15	60	57.1	214	2	S76379	hypothetical protein	3.36e+00
16	60	57.1	598	2	D69292	aldehyde ferredoxin o	3.36e+00
17	60	57.1	790	2	G02676	cadherin-14 - human	3.36e+00
18	59	56.2	154	1	B64467	conserved hypothetical	5.06e+00
19	59	56.2	245	2	F64465	hypothetical protein	5.06e+00
20	59	56.2	452	2	A75213	tldd protein PAB0154	5.06e+00
21	59	56.2	624	2	T00044	vacuolar sorting rece	7.59e+00
22	58	55.2	270	2	G69469	conserved hypothetical	7.59e+00
23	58	55.2	593	2	T04446	hypothetical protein	7.59e+00

24	58	55.2	1069	2	T00040	BH-protocadherin PCDH	7.59e-00
25 <td>58<td>55.2<td>1072<td>2<th>T00041</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td></td></td>	58 <td>55.2<td>1072<td>2<th>T00041</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td></td>	55.2 <td>1072<td>2<th>T00041</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td>	1072 <td>2<th>T00041</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td>	2 <th>T00041</th> <th>BH-protocadherin PCDH</th> <th>7.59e-00</th>	T00041	BH-protocadherin PCDH	7.59e-00
26 <td>58<td>55.2<td>1200<td>2<th>T00042</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td></td></td>	58 <td>55.2<td>1200<td>2<th>T00042</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td></td>	55.2 <td>1200<td>2<th>T00042</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td></td>	1200 <td>2<th>T00042</th><th>BH-protocadherin PCDH</th><th>7.59e-00</th></td>	2 <th>T00042</th> <th>BH-protocadherin PCDH</th> <th>7.59e-00</th>	T00042	BH-protocadherin PCDH	7.59e-00
27 <td>58<td>55.2<td>3027<td>2<th>J01917</th><th>polyprotein - parsnip</th><th>7.59e-00</th></td></td></td></td>	58 <td>55.2<td>3027<td>2<th>J01917</th><th>polyprotein - parsnip</th><th>7.59e-00</th></td></td></td>	55.2 <td>3027<td>2<th>J01917</th><th>polyprotein - parsnip</th><th>7.59e-00</th></td></td>	3027 <td>2<th>J01917</th><th>polyprotein - parsnip</th><th>7.59e-00</th></td>	2 <th>J01917</th> <th>polyprotein - parsnip</th> <th>7.59e-00</th>	J01917	polyprotein - parsnip	7.59e-00
28 <td>57<td>54.3<td>209<td>2<th>C64685</th><th>ribonuclease H1 - He</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>209<td>2<th>C64685</th><th>ribonuclease H1 - He</th><th>1.13e-01</th></td></td></td>	54.3 <td>209<td>2<th>C64685</th><th>ribonuclease H1 - He</th><th>1.13e-01</th></td></td>	209 <td>2<th>C64685</th><th>ribonuclease H1 - He</th><th>1.13e-01</th></td>	2 <th>C64685</th> <th>ribonuclease H1 - He</th> <th>1.13e-01</th>	C64685	ribonuclease H1 - He	1.13e-01
29 <td>57<td>54.3<td>209<td>2<th>F71830</th><th>ribonuclease h1 - He</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>209<td>2<th>F71830</th><th>ribonuclease h1 - He</th><th>1.13e-01</th></td></td></td>	54.3 <td>209<td>2<th>F71830</th><th>ribonuclease h1 - He</th><th>1.13e-01</th></td></td>	209 <td>2<th>F71830</th><th>ribonuclease h1 - He</th><th>1.13e-01</th></td>	2 <th>F71830</th> <th>ribonuclease h1 - He</th> <th>1.13e-01</th>	F71830	ribonuclease h1 - He	1.13e-01
30 <td>57<td>54.3<td>504<td>1<th>G71248</th><th>tldd homolog PH0246 -</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>504<td>1<th>G71248</th><th>tldd homolog PH0246 -</th><th>1.13e-01</th></td></td></td>	54.3 <td>504<td>1<th>G71248</th><th>tldd homolog PH0246 -</th><th>1.13e-01</th></td></td>	504 <td>1<th>G71248</th><th>tldd homolog PH0246 -</th><th>1.13e-01</th></td>	1 <th>G71248</th> <th>tldd homolog PH0246 -</th> <th>1.13e-01</th>	G71248	tldd homolog PH0246 -	1.13e-01
31 <td>57<td>54.3<td>796<td>2<th>I48277</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>796<td>2<th>I48277</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td></td>	54.3 <td>796<td>2<th>I48277</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td>	796 <td>2<th>I48277</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td>	2 <th>I48277</th> <th>cadherin-11 - mouse</th> <th>1.13e-01</th>	I48277	cadherin-11 - mouse	1.13e-01
32 <td>57<td>54.3<td>796<td>2<th>I48556</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>796<td>2<th>I48556</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td></td>	54.3 <td>796<td>2<th>I48556</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td></td>	796 <td>2<th>I48556</th><th>cadherin-11 - mouse</th><th>1.13e-01</th></td>	2 <th>I48556</th> <th>cadherin-11 - mouse</th> <th>1.13e-01</th>	I48556	cadherin-11 - mouse	1.13e-01
33 <td>57<td>54.3<td>796<td>2<th>A53584</th><th>OB-cadherin precursor</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>796<td>2<th>A53584</th><th>OB-cadherin precursor</th><th>1.13e-01</th></td></td></td>	54.3 <td>796<td>2<th>A53584</th><th>OB-cadherin precursor</th><th>1.13e-01</th></td></td>	796 <td>2<th>A53584</th><th>OB-cadherin precursor</th><th>1.13e-01</th></td>	2 <th>A53584</th> <th>OB-cadherin precursor</th> <th>1.13e-01</th>	A53584	OB-cadherin precursor	1.13e-01
34 <td>57<td>54.3<td>796<td>2<th>A38992</th><th>cadherin 11 precursor</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>796<td>2<th>A38992</th><th>cadherin 11 precursor</th><th>1.13e-01</th></td></td></td>	54.3 <td>796<td>2<th>A38992</th><th>cadherin 11 precursor</th><th>1.13e-01</th></td></td>	796 <td>2<th>A38992</th><th>cadherin 11 precursor</th><th>1.13e-01</th></td>	2 <th>A38992</th> <th>cadherin 11 precursor</th> <th>1.13e-01</th>	A38992	cadherin 11 precursor	1.13e-01
35 <td>57<td>54.3<td>798<td>2<th>S62791</th><th>probable lipoprotein</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>798<td>2<th>S62791</th><th>probable lipoprotein</th><th>1.13e-01</th></td></td></td>	54.3 <td>798<td>2<th>S62791</th><th>probable lipoprotein</th><th>1.13e-01</th></td></td>	798 <td>2<th>S62791</th><th>probable lipoprotein</th><th>1.13e-01</th></td>	2 <th>S62791</th> <th>probable lipoprotein</th> <th>1.13e-01</th>	S62791	probable lipoprotein	1.13e-01
36 <td>57<td>54.3<td>887<td>1<th>I0CHCL</th><th>E-cadherin precursor</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>887<td>1<th>I0CHCL</th><th>E-cadherin precursor</th><th>1.13e-01</th></td></td></td>	54.3 <td>887<td>1<th>I0CHCL</th><th>E-cadherin precursor</th><th>1.13e-01</th></td></td>	887 <td>1<th>I0CHCL</th><th>E-cadherin precursor</th><th>1.13e-01</th></td>	1 <th>I0CHCL</th> <th>E-cadherin precursor</th> <th>1.13e-01</th>	I0CHCL	E-cadherin precursor	1.13e-01
37 <td>57<td>54.3<td>1400<td>1<th>I38185</th><th>protein-tyrosine kina</th><th>1.13e-01</th></td></td></td></td>	57 <td>54.3<td>1400<td>1<th>I38185</th><th>protein-tyrosine kina</th><th>1.13e-01</th></td></td></td>	54.3 <td>1400<td>1<th>I38185</th><th>protein-tyrosine kina</th><th>1.13e-01</th></td></td>	1400 <td>1<th>I38185</th><th>protein-tyrosine kina</th><th>1.13e-01</th></td>	1 <th>I38185</th> <th>protein-tyrosine kina</th> <th>1.13e-01</th>	I38185	protein-tyrosine kina	1.13e-01
38 <td>56<td>53.3<td>423<td>2<th>T06774</th><th>cell division protein</th><th>1.68e-01</th></td></td></td></td>	56 <td>53.3<td>423<td>2<th>T06774</th><th>cell division protein</th><th>1.68e-01</th></td></td></td>	53.3 <td>423<td>2<th>T06774</th><th>cell division protein</th><th>1.68e-01</th></td></td>	423 <td>2<th>T06774</th><th>cell division protein</th><th>1.68e-01</th></td>	2 <th>T06774</th> <th>cell division protein</th> <th>1.68e-01</th>	T06774	cell division protein	1.68e-01
39 <td>56<td>53.3<td>713<td>2<th>B38992</th><th>cadherin 13 precursor</th><th>1.68e-01</th></td></td></td></td>	56 <td>53.3<td>713<td>2<th>B38992</th><th>cadherin 13 precursor</th><th>1.68e-01</th></td></td></td>	53.3 <td>713<td>2<th>B38992</th><th>cadherin 13 precursor</th><th>1.68e-01</th></td></td>	713 <td>2<th>B38992</th><th>cadherin 13 precursor</th><th>1.68e-01</th></td>	2 <th>B38992</th> <th>cadherin 13 precursor</th> <th>1.68e-01</th>	B38992	cadherin 13 precursor	1.68e-01
40 <td>56<td>53.3<td>756<td>2<th>S67433</th><th>hypothetical protein</th><th>1.68e-01</th></td></td></td></td>	56 <td>53.3<td>756<td>2<th>S67433</th><th>hypothetical protein</th><th>1.68e-01</th></td></td></td>	53.3 <td>756<td>2<th>S67433</th><th>hypothetical protein</th><th>1.68e-01</th></td></td>	756 <td>2<th>S67433</th><th>hypothetical protein</th><th>1.68e-01</th></td>	2 <th>S67433</th> <th>hypothetical protein</th> <th>1.68e-01</th>	S67433	hypothetical protein	1.68e-01
41 <td>55<td>52.4<td>211<td>2<th>F69391</th><th>hypothetical protein</th><th>2.48e+01</th></td></td></td></td>	55 <td>52.4<td>211<td>2<th>F69391</th><th>hypothetical protein</th><th>2.48e+01</th></td></td></td>	52.4 <td>211<td>2<th>F69391</th><th>hypothetical protein</th><th>2.48e+01</th></td></td>	211 <td>2<th>F69391</th><th>hypothetical protein</th><th>2.48e+01</th></td>	2 <th>F69391</th> <th>hypothetical protein</th> <th>2.48e+01</th>	F69391	hypothetical protein	2.48e+01
42 <td>55<td>52.4<td>591<td>2<th>H72474</th><th>probable acylamino-ac</th><th>2.48e+01</th></td></td></td></td>	55 <td>52.4<td>591<td>2<th>H72474</th><th>probable acylamino-ac</th><th>2.48e+01</th></td></td></td>	52.4 <td>591<td>2<th>H72474</th><th>probable acylamino-ac</th><th>2.48e+01</th></td></td>	591 <td>2<th>H72474</th><th>probable acylamino-ac</th><th>2.48e+01</th></td>	2 <th>H72474</th> <th>probable acylamino-ac</th> <th>2.48e+01</th>	H72474	probable acylamino-ac	2.48e+01
43 <td>55<td>52.4<td>882<td>1<th>H72474</th><th>cadherin 1 precursor</th><th>2.48e+01</th></td></td></td></td>	55 <td>52.4<td>882<td>1<th>H72474</th><th>cadherin 1 precursor</th><th>2.48e+01</th></td></td></td>	52.4 <td>882<td>1<th>H72474</th><th>cadherin 1 precursor</th><th>2.48e+01</th></td></td>	882 <td>1<th>H72474</th><th>cadherin 1 precursor</th><th>2.48e+01</th></td>	1 <th>H72474</th> <th>cadherin 1 precursor</th> <th>2.48e+01</th>	H72474	cadherin 1 precursor	2.48e+01
44 <td>55<td>52.4<td>1196<td>2<th>S46430</th><th>botulinum neurotoxin-</th><th>2.48e+01</th></td></td></td></td>	55 <td>52.4<td>1196<td>2<th>S46430</th><th>botulinum neurotoxin-</th><th>2.48e+01</th></td></td></td>	52.4 <td>1196<td>2<th>S46430</th><th>botulinum neurotoxin-</th><th>2.48e+01</th></td></td>	1196 <td>2<th>S46430</th><th>botulinum neurotoxin-</th><th>2.48e+01</th></td>	2 <th>S46430</th> <th>botulinum neurotoxin-</th> <th>2.48e+01</th>	S46430	botulinum neurotoxin-	2.48e+01
45 <td>55<td>52.4<td>1196<td>2<th>J01467</th><th>toxin, nontoxic compo</th><th>2.48e+01</th></td></td></td></td>	55 <td>52.4<td>1196<td>2<th>J01467</th><th>toxin, nontoxic compo</th><th>2.48e+01</th></td></td></td>	52.4 <td>1196<td>2<th>J01467</th><th>toxin, nontoxic compo</th><th>2.48e+01</th></td></td>	1196 <td>2<th>J01467</th><th>toxin, nontoxic compo</th><th>2.48e+01</th></td>	2 <th>J01467</th> <th>toxin, nontoxic compo</th> <th>2.48e+01</th>	J01467	toxin, nontoxic compo	2.48e+01

ALIGNMENTS

RESULT 1
ENTRY IUHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS A41088
REFERENCE A41088
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references GB:M76482; NRD:9190751; PIDN:AAA60230.1; PID:9190752
#molecule_type mRNA
#accession A41088
#residues 1-999 #label AMA
#cross-references GB:M76482; NRD:9190751; PIDN:AAA60230.1; PID:9190752

GENETICS

#gene GDB:DSG3
#cross-references GDB:134030; OMIM:169615
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE 1-23
24-49
50-999
50-615 #domain signal sequence #status predicted #label SIG
123 #domain propeptide #status predicted #label PRO
24-49 #product desmoglein homolog #status predicted #label
50-999 MAT
50-615 #domain extracellular #status predicted #label EXT
123 #domain cadherin repeat homology #label CR1
160-267 #domain cadherin repeat homology #label CR2
270-383 #domain cadherin repeat homology #label CR3
390-495 #domain cadherin repeat homology #label CR4
496-598 #domain cadherin repeat homology #label CR5
616-639 #domain transmembrane #status predicted #label TMN
640-999 #domain intracellular #status predicted #label INT
910-938 #domain desmoglein repeat #label DGL
937-966 #domain desmoglein repeat #label DGL
110,180,545 #binding_site carbohydrate (asn) (covalent) #status predicted

SUMMARY #length 999 #molecular_weight 107502 #checksum 8311

Query Match 100.0%; Score 105; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.33e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 97 FGIFVVDKNTGDINI 111
 OY 1 FGIFVVDKNTGDINI 15

RESULT 2
 ENTRY 1JB0G1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S14603: A38872; A37785; S38721; A48173; S24412
 S14603

REFERENCE
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submision submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 ##molecule_type mRNA
 ##residues 1-1043 #label KOC
 ##cross-references EMBL:X58466; NID:G306; PIDN:CAA41380.1; PID:G307

REFERENCE A38872
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 ##molecule_type mRNA
 ##residues 1-87;968-1043 #label KO2
 ##cross-references GB:S64268; GB:S64270

REFERENCE A37785
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 ##molecule_type mRNA
 ##residues 44-123; 'V', 125-493 #label GCO
 ##cross-references GB:X58165; NID:G162966; PIDN:AAA62709.1; PID:G552318

REFERENCE S38721
 #authors Zimbelmann, R.
 #submision submitted to the EMBL Data Library, February 1991
 #accession S38721
 ##molecule_type mRNA
 ##residues 44-1043 #label ZIM
 ##cross-references EMBL:X57784; NID:G436061; PIDN:CAA40930.1; PID:G436062

REFERENCE A48173
 #authors Koch, P.J.; Walsh, M.J.; Schmeltz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references MUID:91168965
 #accession A48173
 ##molecule_type mRNA
 ##residues 44-1001; 'AQPASAR' #label KO3
 ##cross-references GB:X57784
 #note this sequence has been revised in references A38872 and S38721

GENETICS DSG1
 #gene superfamily cadherin: cadherin repeat homology
 #CLASSIFICATION calcium binding: cell adhesion: duplication: glycoprotein:
 KEYWORDS

FEATURE 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 160,496

transmembrane protein

SUMMARY #length 1043 #molecular_weight 112242 #checksum 6897
 Query Match 80.0%; Score 84; DB 1; Length 1043;
 Best Local Similarity 60.0%; Pred. No. 6,14e-05;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 FGIFVVDKNTGDINI 111
 OY 1 FGIFVVDKNTGDINI 15

RESULT 3
 ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 22-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 30-Jun-1993

ACCESSIONS S16906; A39706; A61254; A61279; S16158
 S16906

REFERENCE
 #authors Buxton, R.S.
 #submision submitted to the EMBL Data Library, November 1990
 #accession S16906
 ##molecule_type mRNA
 ##residues 1-1049 #label BUX
 ##cross-references EMBL:X56554; NID:G30505; PIDN:CAA39976.1; PID:G30506

REFERENCE A39706
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Attaliotis, P.; Poynter, D.; Arneemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 ##molecule_type mRNA
 ##residues 24-1049 #label WHE
 ##cross-references GB:X56554

REFERENCE A61254
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references MUID:92121251
 #accession A61254
 ##molecule_type mRNA
 ##residues 26-1049 #label NIT

REFERENCE A61279
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.

```

GENERIC# #cross-references GB:S64273
#gene GDB:DSG2
#cross-references GDB:128808; OMIM:125671
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
          membrane protein
FEATURE
51-128 #domain cadherin repeat homology #label CR1\
161-271 #domain cadherin repeat homology #label CR2
SUMMARY #length 1117 #molecular_weight 122384 #checksum 7660

Query Match 78.1% Score 82; DB 2; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1.63e-04;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0
Oy 1 FGIFVFNKDTGELNY 112
      |||||:||||:|
      1 FGIFVFNKDTGDIINI 15

Db 98 FGIFVFNKDTGELNY 112
      |||||:||||:|
      1 FGIFVFNKDTGDIINI 15

RESULT 5
ENTRY IJCODE #type complete
TITLE desmocollin 1a - bovine
ALTERNATE_NAMES desmocollin BDCM; desmosomal glycoprotein 2
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999

ACCESSIONS A43838; B43838; A38456; A60714; S14542
REFERENCE A43838
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimhmann, R.;
#journal Schmelz, M.; Franke, W.W.
#title Differentiation (1991) 47:29-36
Amino acid sequence of bovine muzzle epithelial desmocollin
derived from cloned cDNA: a novel subtype of desmosomal
cadherins.
#cross-references MUID:9208912
#accession A43838
#molecule_type mRNA
#residues 1-761 #label KOX
#cross-references GB:S58029; GB:S57985; NID:g453583; PIDN:CAA41088.1;
PID:g453584

#accession B43838
#molecule_type protein
#residues 1-32;65-76;148-159;164-176;190-205;208-219;238-256;
361-375;377-388;478-486 #label KO2
#experimental_source muzzle epithelium
#note #sequence extracted from NCBI backbone

REFERENCE A38456
#authors Collins, J.E.; Legan, P.K.; Kenny, T.P.; MacGarvie, J.;
#journal Holton, J.L.; Garrod, D.R.
#title J. Cell Biol. (1991) 113:381-391
Cloning and sequence analysis of desmosomal glycoproteins 2
and 3 (desmocollins): cadherin-like desmosomal adhesion
molecules with heterogeneous cytoplasmic domains.
#cross-references MUID:91185414
#accession A38456
#molecule_type mRNA
#residues 606-761 #label COL
#cross-references EMBL:X56967; NID:g310; PIDN:CAA0287.1; PID:g311
REFERENCE A60714
#authors Holton, J.L.; Kenny, T.P.; Legan, P.K.; Collins, J.E.; Keen
#journal J.N.; Sharma, R.; Garrod, D.R.
#title J. Cell Sci. (1990) 97:239-246
Desmosomal glycoproteins 2 and 3 (desmocollins) show
N-terminal similarity to calcium-dependent cell-cell
adhesion molecules.
#cross-references MUID:9115997
#accession A60714
#molecule_type protein
#residues 1-6;'A','8-9','R','11-17','RCE','21-23 #label HOL
#experimental_source nasal epithelium

```

```
GENETICS
#gene
CLASSIFICATION
KEYWORDS
#superfamily cadherin; cadherin repeat homology
#alternative splicing; calcium binding; cell adhesion;
#duplication; glycoprotein; phosphoprotein; transmembrane
protein

FEATURE
1-761
1-561
3-108
11-220
223-338
339-444
445-561
562-582
583-761
31,265,413
584,588,678
585
605
671
681
682
SUMMARY
#length 761 #molecular-weight 85170 #checksum 3508

Query Match
Best Local Similarity 53.8%; Score 66; DB 1; Length 761;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 48 FNLFIEKDTGDI 60
1 : : : : :
1 FGIFVVDKNTGDI 13

RESULT 6
ENTRY
TITLE 150180 #type complete
ORGANISM cadherin-7 - chicken
ID 21-Feb-1997 #sequence-revision 21-Feb-1997 #text-change
DATE 20-Aug-1999
ACCESSIONS 150180
REFERENCE 150178
#authors Nakagawa, S.; Takeichi, M.
#journal Development (1995) 121:1321-1332
#title Neural crest cell-cell adhesion controlled by sequential and
#subpopulation-specific expression of novel cadherins.
#cross-references NUID:95309115
#accession 150180
#status preliminary; translated from GB/EMBL/DBJ
#molecule-type mRNA
#residues 1-785 #label MAK
#cross-references GB:042150; NID:9868000; PIDN:BA07721.1; PID:9868001
CLASSIFICATION #superfamily cadherin repeat homology
SUMMARY #length 785 #molecular-weight 87171 #checksum 8490

Query Match
Best Local Similarity 62.9%; Score 66; DB 2; Length 785;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 SIFIIDENTGDIH 105
1 : : : : :
2 FGIFVVDKNTGDI 14

RESULT 7
ENTRY IJBDF #type complete
TITLE desmocollin 1b precursor - bovine
ALTERNATE_NAMES desmosomal glycoprotein 3

GENETICS
#gene
CLASSIFICATION
KEYWORDS
#formal_name Bos primigenius taurus #common_name cattle
#alternative splicing; calcium binding; cell adhesion;
#duplication; glycoprotein; phosphoprotein; transmembrane
protein

FEATURE
1-29
1-132
133-839
133-682
135-240
243-352
355-470
471-576
577-682
694-714
716-839
163,398,545
716,720,810
737
803,830
813
814
SUMMARY
#length 839 #molecular-weight 93521 #checksum 6365

Query Match
Best Local Similarity 53.8%; Score 66; DB 1; Length 839;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 180 FNLFIEKDTGDI 192
1 : : : : :
1 FGIFVVDKNTGDI 13

RESULT 8
ENTRY B48910 #type fragment
TITLE desmocollin 1b precursor - human (fragment)
```

ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
13-Aug-1997

ACCESSIONS B48910
REFERENCE #authors King, I.A.; Arneemann, J.; Spurr, N.K.; Buxton, R.S.
#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1 desmocollin and its assignment to chromosome 18.

#cross-references MUID:94116981
#accession B48910
#status Preliminary
#molecule_type mRNA
#residues 1-770 #label KIN
#cross-references GB:X72929

GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
KEYWORDS #alternative_splicing: calcium binding; cell adhesion; glycoprotein; phosphoprotein; transmembrane protein

FEATURE 67-172
#domain_cadherin repeat homology #label CR1
SUMMARY #length 770 #checksum 4861

Query Match 61.9%: Score 65; DB 2; Length 770;
Best Local Similarity 53.8%: Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 112 FNLFIKDTGDI 124
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 9
ENTRY A48910 #type fragment
TITLE desmocollin 1a precursor - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
15-Aug-1997

ACCESSIONS A48910
REFERENCE #authors King, I.A.; Arneemann, J.; Spurr, N.K.; Buxton, R.S.
#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1 desmocollin and its assignment to chromosome 18.

#cross-references MUID:94116981
#accession A48910
#status Preliminary
#molecule_type mRNA
#residues 1-824 #label KIN
#cross-references GB:X72925

GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
KEYWORDS #alternative_splicing: calcium binding; cell adhesion; glycoprotein; phosphoprotein; transmembrane protein

FEATURE 67-172
#domain_cadherin repeat homology #label CR1
SUMMARY #length 824 #checksum 2645

Query Match 61.9%: Score 65; DB 2; Length 824;
Best Local Similarity 53.8%: Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 112 FNLFIKDTGDI 124
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 10
ENTRY I37281 #type complete
TITLE Dscia precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
20-Aug-1999

ACCESSIONS I37281
REFERENCE #authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified epithelia.

#cross-references MUID:93283249
#accession I37281
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-840 #label RES
#cross-references EMBL:234532; NID:g505536; PIDN:CAA84278.1; PID:g505537

CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
SUMMARY #length 840 #molecular_weight 93848 #checksum 3174

Query Match 61.9%: Score 65; DB 2; Length 840;
Best Local Similarity 53.8%: Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 182 FNLFIKDTGDI 194
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 11
ENTRY I37282 #type complete
TITLE Dscib precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
20-Aug-1999

ACCESSIONS I37282
REFERENCE #authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified epithelia.

#cross-references MUID:93283249
#accession I37282
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-894 #label RES
#cross-references EMBL:234532; NID:g505536; PIDN:CAA84279.1; PID:g505538

GENETICS
#introns 829/2
CLASSIFICATION #superfamily_cadherin: cadherin repeat homology
SUMMARY #length 894 #molecular_weight 100044 #checksum 7946

Query Match 61.9%: Score 65; DB 2; Length 894;
Best Local Similarity 53.8%: Pred. No. 4.08e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 182 FNLFIKDTGDI 194
1 : : : : :
QY 1 FGIFVVDKNTGDI 13

RESULT 12
ENTRY S5969 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform beta - Chinese hamster
ALTERNATE_NAMES DNA topoisomerase II isoform beta; DNA-gyrase
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE 15-Feb-1996 #sequence_revision 13-Mar-1997 #text_change
16-Jul-1999

ACCESSIONS S5969; S54154

REFERENCE S59966
#authors Derudder, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.
#journal Biochim. Biophys. Acta (1995) 1264:178-182
#title Cloning and characterization of full-length cDNAs coding for the DNA topoisomerase II beta from Chinese hamster lung cells sensitive and resistant to 9-OH-ellipticine.
#cross-references EMBL:96085121
#accession S59966
#status translation not shown
#molecule-type mRNA
#residues 1-1612 ##label DER
#cross-references EMBL:X86455; NID:9790987; PIDN:CAA60173.1; PID:9790988

##experimental_source lung
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase: phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS ATP: DNA binding; DNA replication; isomerase; nucleus
FEATURE
697-927 #domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
#length 1612 #molecular-weight 182074 #checksum 5332

SUMMARY
Query Match 58.1%; Score 61; DB 2; Length 1612;
Best Local Similarity 72.7%; Pred. No. 2,22e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 922 IFVDRNTVEI 932
|||||:|:
3 IFVDRNTGDI 13

RESULT 13
ENTRY A39242 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) beta,
splice form 2 - human
ALTERNATE_NAMES DNA topoisomerase II isoform beta-2
CONTAINS DNA topoisomerase II isoform beta-1
ORGANISM #formal name Homo sapiens #common name man
DATE 04-Oct-1991 #sequence-revision 03-May-1996 #text-change 16-Jul-1999

ACCESSIONS S26730; A39242; S10710; S33970; S30191; S41641; S30190
REFERENCE S26730
#authors Jenkins, J.R.; Aylon, P.; Jones, T.; Davies, S.L.; Simmons, D.L.; Harris, A.L.; Sheer, D.; Hickson, I.D.
#journal Nucleic Acids Res. (1992) 20:5587-5592
#title Isolation of cDNA clones encoding the beta isozyme of human DNA topoisomerase II and localisation of the gene to chromosome 3p24.
#cross-references MUID:93087165

SUMMARY
Query Match 58.1%; Score 61; DB 2; Length 1626;
Best Local Similarity 72.7%; Pred. No. 2,22e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 934 IFVDRNTVEI 944
|||||:|:
3 IFVDRNTGDI 13

RESULT 14
ENTRY A65141 #type complete
TITLE hypothetical 10.6 kb protein in gntR-gvt intergenic region -
#formal name Escherichia coli (strain K-12)
ORGANISM Escherichia coli
DATE 12-Sep-1997 #sequence-revision 17-Sep-1997 #text-change 14-Nov-1997

ACCESSIONS A65141
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462

##cross-references GB:X53662; GB:S56813; NID:938324; PIDN:CAA37706.1; PID:938325

REFERENCE S30190
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#journal Biochim. Biophys. Acta (1993) 1172:283-291
#title Novel Hela topoisomerase II is the II-beta isoform: complete coding sequence and homology with other type II topoisomerases.
#cross-references MUID:93192319
#accession S33970
#molecule-type mRNA
#residues 1-23,29-1610,'A',1612-1626 ##label A02
#cross-references EMBL:215111

REFERENCE S30191
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#submission submitted to the EMBL Data Library, September 1992
#accession S30191
#molecule-type mRNA
#residues 596-1430,'S',1432-1610,'A',1612-1626 ##label A01
#cross-references EMBL:215115; NID:9288564; PIDN:CAA78821.1; PID:9288565

REFERENCE S41641
#authors Davies, S.L.; Jenkins, J.R.; Hickson, I.D.
#journal Nucleic Acids Res. (1993) 21:3719-3723
#title Human cells express two differentially spliced forms of topoisomerase II-beta mRNA.
#cross-references MUID:93376494
#accession S41641
#molecule-type DNA
#residues 24-80 ##label DAV
#cross-references EMBL:X71911
#note this sequence represents a long minor splice form, designated beta-2

GENETICS
#gene GDB:TOP2B
#cross-references GDB:131575; OMIM:126431
#map_position 3p24-3p24
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase: phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS alternative splicing; ATP; dimer; isomerase; nucleus
FEATURE
1-1626 #product DNA topoisomerase II beta-2 #status predicted
#label MINR
#product DNA topoisomerase II beta-1 #status predicted
#label MAJR
#domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
#length 1626 #molecular-weight 183296 #checksum 3565

SUMMARY
Query Match 58.1%; Score 61; DB 2; Length 1626;
Best Local Similarity 72.7%; Pred. No. 2,22e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 934 IFVDRNTVEI 944
|||||:|:
3 IFVDRNTGDI 13

RESULT 14
ENTRY A65141 #type complete
TITLE hypothetical 10.6 kb protein in gntR-gvt intergenic region -
#formal name Escherichia coli (strain K-12)
ORGANISM Escherichia coli
DATE 12-Sep-1997 #sequence-revision 17-Sep-1997 #text-change 14-Nov-1997

ACCESSIONS A65141
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462

```

#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97425617
#accession A65141
#status preliminary; nucleic acid sequence not shown;
      translation not shown
##molecule_type DNA
##residues 1-94 #label BLAT
##cross-references GB:AE000421; GB:U00096; NID:g1789854; PID:g1789855;
      UMG:p33446
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene yrb
SUMMARY #length 94 #molecular-weight 10613 #checksum 4152
Query Match 57.1%; Score 60; DB 2; Length 94;
Best Local Similarity 45.5%; Pred. No. 3.36e+00;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
Db 59 FIIDSGEIH 69
|:|:|:|:|:|:|
OY 4 FVVDKNTGDN 14

RESULT 15
ENTRY S76379 #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal name Synecocystis sp.
          PCC 6803
          #variety
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
      21-Aug-1998
ACCESSIONS S76379
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
      Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, E.;
      Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
      Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimpō,
      S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
      Yasuda, M.; Tabata, S.
      DNA Res. (1996) 3:109-136
#journal #title Sequence analysis of the genome of the unicellular
      cyanobacterium Synecocystis sp. PCC6803. II. Sequence
      determination of the entire genome and assignment of
      potential protein-coding regions.
#cross-references MUID:97061201
#accession S76379
#status preliminary
##molecule_type DNA
##residues 1-214 #label KAN
##cross-references EMBL:D64000; GB:AB001339; NID:g1001484; PID:d1010882;
      PID:g1001603
#note the nucleotide sequence was submitted to the EMBL Data
      Library, June 1996
SUMMARY #length 214 #molecular-weight 24539 #checksum 4859
Query Match 57.1%; Score 60; DB 2; Length 214;
Best Local Similarity 42.9%; Pred. No. 3.36e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
Db 49 FGVLMDSTGEIS 62
|:|:|:|:|:|:|
OY 1 FGIFVDKNTGDN 14

```

Search completed: Sat May 13 07:04:57 2000
 Job time : 8 secs.

KM Calcium-binding; Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT SWIMIN 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SO SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 105; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 2,32e-11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 97 FGIFFVNDKNTGDIINI 111
 0Y 1 FGIFFVNDKNTGDIINI 15

RESULT 2
 ID DSG1_BOVIN STANDARD: PRT: 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DSG1).
 OS DSG1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MUSCLE EPITHELIUM;
 RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
 RA Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE=MUSCLE EPITHELIUM;
 RX MEDLINE: 9116965.
 RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
 RA Zimbelmann R., Franke W.W.;
 RT Identification of desmoglein, a constitutive desmosomal
 RT glycoprotein, as a member of the cadherin family of cell adhesion
 RT molecules.*;
 RL Eur. J. Cell Biol. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RX MEDLINE: 92037656.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
 RA Franke W.W.;
 RT Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.*;
 RL Eur. J. Cell Biol. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RX MEDLINE: 91097553.
 RA Goodwin L., Hill J.E., Raynor K., Raazi L., Manabe M., Cowin P.;
 RA Desmoglein shows extensive homology to the cadherin family of cell
 RA adhesion molecules.*;
 RT Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
 CC -I- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -I- TISSUE SPECIFICITY: EPIDERMIS, MIZZLE, TONGUE AND ESOPHAGUS.
 CC -I- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -I- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X58466; CA41380.1; -;
 DR EMBL: X57784; CA440930.1; -;
 DR EMBL: M58165; AAA62709.1; -;
 DR PIR: S14603; IJBOGI.
 DR HSSP: P09803; 1EDH.
 DR PROSITE: PS00232; CADHERIN: 2.
 DR PFM: PF00028; cadherin: 3.
 DR Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KM Calcium-binding; Repeat.
 KM SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 1043 DESMOGLEIN 1.
 FT DOMAIN 50 548 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 549 573 POTENTIAL.
 FT DOMAIN 574 1043 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 574 580 CADHERIN 1.
 FT REPEAT 580 588 CADHERIN 2.
 FT REPEAT 588 596 CADHERIN 3.
 FT REPEAT 596 604 CADHERIN 4.
 FT REPEAT 604 612 DESMOGLEIN REPEAT 1.
 FT REPEAT 612 620 DESMOGLEIN REPEAT 2.
 FT REPEAT 620 628 DESMOGLEIN REPEAT 3.
 FT REPEAT 628 636 DESMOGLEIN REPEAT 4.
 FT REPEAT 636 644 DESMOGLEIN REPEAT 5.
 FT CARBOHYD 110 110 GLY/SER-RICH.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 496 496 POTENTIAL.
 FT CONFLICT 124 124 I -> V (IN REF. 4).
 SO SEQUENCE 1043 AA: 112243 MW: ADE46133FB87C11 CRC64;

Query Match 80.0%; Score 84; DB 1; Length 1043;
 Best Local Similarity 60.0%; Pred. No. 4,29e-06;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVNDKNTGDIINI 111
 0Y 1 FGIFFVNDKNTGDIINI 15

RESULT 3
 ID DSG1_HUMAN STANDARD: PRT: 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DSG1).
 OS DSG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KERATINOCYTES;
 RX MEDLINE: 91271279.
 RA Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Arnmann J., Rutan A.J., Pidsley S.C., Watt F.M., Rees D.A.,
 RA Buxton R.S., Magee A.I.;
 RT Desmosomal glycoprotein Dsg1, a component of intercellular desmosome

RT junctions, is related to the cadherin family of cell adhesion
RT molecules.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONILAND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X56654; CAA39976.1; -
DR PIR: S16906; IJHUG1.
DR HSSP: P09803; IEDH.
DR MIM: 125670; -
DR PROSITE: PS00232; CADHERIN. 2.
DR PFAM: PF00028; cadherin; 4.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1049
FT DOMAIN 50 545
FT TRANSMEM 546 570
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 813 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SO SEQUENCE 1049 AA; 113715 MW; EBE12565589D6619 CRC64;

Query Match 80.0%; Score 84; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 4.29e-06;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVINKTGEINI 111
OY 1 FGIFVVKNTGDINI 15

RESULT 4
ID DSG2_HUMAN STANDARD; PRT; 1117 AA.
AC 014126;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 2 PRECURSOR (HDGC).
GN DSG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-COLON CARCINOMA:
RX MEDLINE; 94192736.

RA Schaefer S., Koch P.J., Franke W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT expression catalogue of the desmoglein subfamily of desmosomal
RT cadherins.";
RL Exp. Cell Res. 211:391-399(1994).
RN [2]
RN SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE; 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Franke W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene.";
RL Eur. J. Cell Biol. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z26317; CAA81226.1; -
DR HSSP: P15116; INCI.
DR MIM: 125671; -
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PFAM: PF00028; cadherin; 4.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW Cytoskeleton; Calcium-binding.
FT SIGNAL 1 23
FT PROPEP 24 48
FT CHAIN 49 1117
FT DOMAIN 49 608
FT TRANSMEM 609 633
FT DOMAIN 634 1117
FT REPEAT 49 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 880 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SO SEQUENCE 1117 AA; 122385 MW; 223857FED70B289 CRC64;

Query Match 78.1%; Score 82; DB 1; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1.28e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 98 FGIFVFKNGELNV 112
OY 1 FGIFVVKNTGDINI 15

RESULT 5
ID DSC1_BOVIN STANDARD; PRT; 893 AA.
AC 00107; 028095;
DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, last sequence update)
 DE 15-FEB-2000 (Rel. 39, last annotation update)
 DT DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSCI
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RC TISSUE-EPIDERMIS;
 RX MEDLINE: 91185414.
 RA Collins J.E., Legan P.K., Kenny T.P., Macgavyle J., Holton J.L.,
 Garrod D.R.;
 RT "Cloning and sequence analysis of desmosomal glycoproteins 2 and 3
 RT (desmocollins): cadherin-like desmosomal adhesion molecules with
 RT heterogeneous cytoplasmic domains.";
 RL J. Cell Biol. 113:381-391(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RX MEDLINE: 91239591.
 RA Mechanic S., Raynor K., Hill J.E., Cowin P.;
 RT "Desmocollins form a distinct subset of the cadherin family of cell
 RT adhesion molecules.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4476-4480(1991).
 RN [3]
 RP SEQUENCE OF 133-893 FROM N.A. (1A), AND PARTIAL SEQUENCE.
 RC TISSUE-MUZZLE EPITHELIUM;
 RX MEDLINE: 92008912.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Schmelz M.,
 Franke W.W.;
 RT Amino acid sequence of bovine muzzle epithelial desmocollin derived
 RT from cloned cDNA: a novel subtype of desmosomal cadherins.";
 RL Differentiation 47:29-36(1991).
 RN [4]
 RP SEQUENCE OF 133-155.
 RX MEDLINE: 91115997.
 RA Holton J.L., Kenny T.P., Legan P.K., Collins J.E., Keen J.N.,
 Sharma R., Garrod D.R.;
 RT "Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal
 RT similarity to calcium-dependent cell-cell adhesion molecules.";
 RL J. Cell Sci. 97:239-246(1990).
 RN [5]
 RP PHOSPHORYLATION.
 RX MEDLINE: 91009551.
 RA Parrish E.P., Marston J.E., Matley D.L., Measures H.R., Vennart R.,
 Garrod D.R.;
 RT "Size heterogeneity, phosphorylation and transmembrane organisation
 RT of desmosomal glycoproteins 2 and 3 (desmocollins) in MDCK cells.";
 RL J. Cell Sci. 96:239-248(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: 1A/DG2 (SHOWN HERE) AND 1B/DG3;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS AND WEAKLY IN TONGUE PAPILLAE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- PTM: FORM 1A IS PHOSPHORYLATED ON A SERINE BUT FORM 1B IS NOT.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).

DR EMBL: M67489; AAA30492.1; -
 DR EMBL: X56966; CAA40286.1; -
 DR EMBL: X56967; CAA40287.1; -
 DR EMBL: X56968; CAA40289.1; -
 DR EMBL: X56968; CAA40288.1; -
 DR EMBL: X58029; CAA41088.1; -
 DR PIR: B18456; IJBDE.
 DR PIR: A43838; IJBDE.
 DR PIR: A39377; A39377.
 DR HSP: P09803; IEDH.
 DR PRINTS: P00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PFAM: PF00028; cadherin; 5.
 KM Cell adhesion: Glycoprotein: Phosphorylation: Transmembrane: Signal;
 KM Repeat: Cytoskeleton: Calcium-binding; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPE 30 132
 FT CHAIN 133 893
 FT DOMAIN 133 692
 FT TRANSMEM 693 715
 FT DOMAIN 716 893
 FT REPEAT 133 240
 FT REPEAT 241 352
 FT REPEAT 353 470
 FT REPEAT 471 574
 FT REPEAT 575 682
 FT CARBOHYD 163 163
 FT CARBOHYD 398 398
 FT CARBOHYD 545 545
 FT VARIANT 519 519
 FT VARIANT 788 788
 FT VARSPLIC 829 839
 FT VARSPLIC 840 893
 FT CONFLICT 485 485
 FT SEQUENCE 893 AA; 99647 MW; A45A4DB3051FC9 CRC64;
 SO
 Query Match 62.9%; Score 66; DB 1; Length 893;
 Best Local Similarity 53.8%; Pred. No. 5,19e-02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 180 FNLFIEKDTGDI 192
 Qy 1 FGIFVVDKNTGDI 13
 RESULT 6
 ID DSCI_HUMAN STANDARD: PRT: 894 AA.
 AC 008554;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSCI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORSKIN;
 RX MEDLINE: 93283249.
 RA Theis D.G., Koch P.J., Franke W.W.;
 RT "Differential synthesis of type 1 and type 2 desmocollin mRNAs in
 RT human stratified epithelia.";
 RL Int. J. Dev. Biol. 37:101-110(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORSKIN;
 RX Zimbelmann R.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKIN;

RX MEDLINE: 94116981.
 RA King I.A., Armemann J., Spurr N.K., Buxton R.S.:
 RT "Cloning of the cDNA (DSCL) coding for human type 1 desmocollin and
 its assignment to chromosome 18.";
 RN Genomics 18:185-194(1993).
 RP SEQUENCE OF 135-151 AND 283-292.
 RX MEDLINE: 91323543.
 RA King I.A., Magee A.I., Rees D.A., Buxton R.S.:
 RT "Keratinization is associated with the expression of a new protein
 related to the desmosomal cadherins Dsg1/111.";
 RN FEBS Lett. 286:9-12(1991).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: 1A/DG2 (SHOWN HERE) AND 1B/DG3.
 CC -1- ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN EPIDERMIS, LESS IN LYMPH
 NODE AND TONGUE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 234522; CAAB4279.1; -
 DR EMBL: 234522; CAAB4278.1; -
 DR EMBL: X72925; CAAS1428.1; -
 DR EMBL: X72925; CAAS1429.1; -
 DR HSSP: P09803; 1EDH.
 DR MIM: 125643; -
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
 KM Cytoskeleton; Calcium-binding; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPEP 30 134
 FT CHAIN 135 884
 FT DOMAIN 135 691
 FT TRANSMEM 692 714
 FT DOMAIN 715 894
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 575
 FT REPEAT 576 682
 FT CARBOHYD 165 165
 FT CARBOHYD 546 546
 FT VARSPLIC 830 840
 FT VARSPLIC 841 894
 FT CONFLICT 132 132
 FT SEQUENCE 894 AA; 100044 MW; 44BA33038699E3E1 CRC64;
 SQ
 Query Match 61.9%; Score 65; DB 1; Length 894;
 Best Local Similarity 53.8%; Pred. No. 8, 44e-02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

RESULT 7
 ID DSCL MOUSE STANDARD: PRT: 886 AA.
 AC P55849;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUL-1998 (rel. 36, Last annotation update)
 DE DESMOCOLLIN 1A/1B PRECURSOR.
 GN DSCL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-SKIN;
 RX MEDLINE: 96420658.
 RA King I.A., O'Brien T.J., Buxton R.S.:
 RT "Expression of the 'skin-type' desmosomal cadherin Dsc1 is closely
 linked to the keratinization of epithelial tissues during mouse
 development.";
 RL J. Invest. Dermatol. 107:531-538(1996).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: FORMS 1A AND 1B ARE PRODUCED BY ALTERNATIVE
 SPLICING OF THE SAME GENE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X97986; CAA66628.1; -
 DR EMBL: X97986; CAA66629.1; -
 DR HSSP: P09803; 1EDH.
 DR MGD: MGI:109173; DSC1.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PFAM: PF00028; cadherin; 5.
 KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KM Calcium-binding; Repeat; Alternative splicing.
 FT SIGNAL 1 29
 FT PROPEP 30 134
 FT CHAIN 135 886
 FT DOMAIN 135 691
 FT TRANSMEM 692 714
 FT DOMAIN 715 886
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 575
 FT REPEAT 576 682
 FT CARBOHYD 130 130
 FT CARBOHYD 165 165
 FT CARBOHYD 546 546
 FT CARBOHYD 613 613
 FT VARSPLIC 822 832
 FT VARSPLIC 840 886
 FT SEQUENCE 886 AA; 98953 MW; F34FB8D578CE92F7 CRC64;
 SQ
 Query Match 58.1%; Score 61; DB 1; Length 886;
 Best Local Similarity 46.2%; Pred. No. 5, 62e-01;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 182 YNLFYKMDTGD 194
 1 FGIVVDKNTGDI 13

RESULT 8
 ID TP2B_MOUSE STANDARD: PRT: 1612 AA.
 AC 064511:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 13-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN
 RP
 RC SEQUENCE FROM N.A.
 RA STRAIN-BALB/C; TISSUE-BRAIN;
 RA Miyake M., Adachi N., Kikuchi A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS.
 CC CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D38046; BAA07236.1; -.
 DR HSSP: P06786; 1BGW.
 DR MGD: MGI:98791; TOP2B.
 DR PRINTS: PRO0418; TP12FAMILY.
 DR PRINTS: PRO0615; CCAATSUBUNTA.
 DR PRINTS: PRO1158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFWAM: PFW0204; DNA_topoisoi1; 1.
 DR KMW: KMW0204; DNA_topoisoi1; 1.
 DR NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 1612 AA; 181863 MW; 331B930065130864 CRC64;
 SQ SEQUENCE 1612 AA; 181863 MW; 331B930065130864 CRC64;
 SO
 Query Match 58.18; Score 61; DB 1; Length 1612;
 Best Local Similarity 72.78; Pred. No. 5.62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 922 IFVVDRTVEI 932
 3 IFVVDKNTGDI 13

RESULT 9
 ID TP2B_CRILLO STANDARD: PRT: 1612 AA.
 AC 064399:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.

OS Cricetus longicaudatus (long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
 RN
 RP
 RC SEQUENCE FROM N.A.
 RA
 RA MEDLINE: 96085121.
 RA Derouille S., Frey S., Delaportie C., Jacquemin-Sablon A.;
 RT Cloning and characterization of full-length cDNAs coding for the DNA
 RT topoisomerase II beta from Chinese hamster lung cells sensitive and
 RT resistant 9-OH-ellipticine.
 RL Biochim. Biophys. Acta 1264:178-182(1995).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II;
 CC MAKES DOUBLE-STRAND BREAKS.
 CC CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X86455; CAA60173.1; -.
 DR HSSP: P06786; 1BGW.
 DR PRINTS: PRO0418; TP12FAMILY.
 DR PRINTS: PRO0615; CCAATSUBUNTA.
 DR PRINTS: PRO1158; TOPISMRASEII.
 DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
 DR PFWAM: PFW0204; DNA_topoisoi1; 1.
 DR KMW: KMW0204; DNA_topoisoi1; 1.
 DR NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;
 SQ SEQUENCE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;
 SO
 Query Match 58.18; Score 61; DB 1; Length 1612;
 Best Local Similarity 72.78; Pred. No. 5.62e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 922 IFVVDRTVEI 932
 3 IFVVDKNTGDI 13

RESULT 10
 ID TP2B_HUMAN STANDARD: PRT: 1626 AA.
 AC 002880:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
 GN TOP2B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN
 RP
 RC SEQUENCE FROM N.A.
 RA MEDLINE: 93087155.
 RA Jenkins J.R., Ayton P., Jones T., Davies S.L., Simmons D.L.,
 RA Harris A.L., Sheer D., Hickson I.D.;
 RT Isolation of cDNA clones encoding the beta isozyme of human DNA
 RT topoisomerase II and localisation of the gene to chromosome 3p24.
 RL Nucleic Acids Res. 20:5587-5592(1992).
 GN

```

RP. SEQUENCE FROM N.A.
RX MEDLINE: 93192319.
RA Austin C.A., Sing J.H., Patel S., Fisher L.M.:
RT "Novel Hela topoisomerase II is the II beta isoform: complete coding
RT sequence and homology with other type II topoisomerases."
RL Biochim. Biophys. Acta 1172:283-291(1993).
RN [3]
RP SEQUENCE OF 1038-1271 FROM N.A.
RX MEDLINE: 90306333.
RA Austin C.A., Fisher L.M.:
RT "Isolation and characterization of a human cDNA clone encoding a
RT novel DNA topoisomerase II homologue from Hela cells."
RL FEBS Lett. 266:115-117(1990).
RN [4]
RP ALTERNATIVE SPLICING.
RX MEDLINE: 93376494.
RA Davies S.L., Jenkins J.R., Hickson I.D.:
RT "Human cells express two differentially spliced forms of
RT topoisomerase II beta mRNA."
RL Nucleic Acids Res. 21:3719-3723(1993).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, CALLED BETA-1 AND BETA-2 ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE TOP2B GENE. THE BETA-2
CC FORM IS SHOWN HERE.
CC -1- PTM: PHOSPHORYLATED.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC .....
DR EMBL: X68060; CAA48197.1; -.
DR EMBL: X71911; -. NOT ANNOTATED_CDS.
DR EMBL: Z15111; CAA78821.1; -.
DR EMBL: Z15115; CAA78821.1; -.
DR EMBL: X53662; CAA37706.1; -.
DR PIR: S26730; S26730.
DR PIR: S10710; S10710.
DR PIR: S41641; S41641.
DR HSP: P06786; IBCW.
DR MIM: 126431.
DR PRINTS: PR00418; TP12FAMILY.
DR PRINTS: PR00615; CCAATSUBUNITA.
DR PRINTS: PR01158; TOPISMRASEII.
DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
DR PFAM: PF00204; DNA_topoisomII; 1.
KW Isomerase: Topoisomerase: DNA-binding; ATP-binding; Phosphorylation;
KW Alternative splicing; Nuclear protein.
FT NP_BIND 182 187 ATP (POTENTIAL).
FT ACT_SITE 826 826 DNA CLEAVAGE (BY SIMILARITY).
FT VARSPIC 24 28 MISSING (IN ISOFORM BETA-1).
FT CONFLICT 1611 1611 T -> A (IN REF. 2).
SQ SEQUENCE 1626 AA; 183296 MW; E7BE9262CC8B04D CRC64;
Query Match 58.1%; Score 61; DB 1; Length 1626;
Best Local Similarity 72.7%; Pred. No. 5,62e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB 934 IFVVDKNTVEI 944
|||||:|:|

```

```

OY 3 IFVVDKNTGDI 13
RESULT 11
ID TP2B_CHICK STANDARD; PRT: 1627 AA.
AC 042131;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN TOP2B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA Nimi A., Harata M., Mizuno S.:
RT Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; GENERALLY LOCATED IN THE
CC NUCLEOLASM.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC .....
DR EMBL: AB007446; BAA2540.1; -.
DR HSP: P06786; IBCW.
DR PRINTS: PR00418; TP12FAMILY.
DR PRINTS: PR00615; CCAATSUBUNITA.
DR PRINTS: PR01158; TOPISMRASEII.
DR PROSITE: PS00177; TOPOISOMERASE-II; 1.
DR PFAM: PF00204; DNA_topoisomII; 1.
KW Isomerase: Topoisomerase: DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 187 192 ATP (POTENTIAL).
FT ACT_SITE 831 831 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 5 23 POLY-GLY.
FT DOMAIN 1265 1268 POLY-LYS.
FT DOMAIN 1388 1391 POLY-ASP.
FT DOMAIN 1393 1396 POLY-ASN.
SQ SEQUENCE 1627 AA; 183245 MW; 8B651D10A2CAD34B CRC64;
Query Match 58.1%; Score 61; DB 1; Length 1627;
Best Local Similarity 72.7%; Pred. No. 5,62e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
DB 939 IFVVDKNTGDI 949
|||||:|:|
OY 3 IFVVDKNTGDI 13
RESULT 12
ID YRHB_ECOLI STANDARD; PRT: 94 AA.
AC P46857;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 10.6 KD PROTEIN IN GNT-R-GGT INTERGENIC REGION (O94).
GN YRHB.
OS Escherichia coli.

```


OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U18997; AAA58244.1; -;
 CC EMBL: AE000421; AAC76471.1; -;
 CC ECGENE: EGI2948; yrb.
 CC KW Hypothetical protein.
 CC SEQUENCE 94 AA: 10613 MW: 8A86265E70E5B06 CRC64;
 SO
 Query Match 57.1%; Score 60; DB 1; Length 94;
 Best Local Similarity 45.5%; Pred. No. 8,91e-01;
 Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;
 DB 59 FIIDKSGEIH 69
 ID 4 FVVDKNTGDIIN 14
 OY
 RESULT 13
 ID CADE_HUMAN STANDARD: PRT: 790 AA.
 AC Q13634;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CADHERIN-14 PRECURSOR.
 GN CDH14.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homiidae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 97184182.
 RX Shibata T., Shimoyama Y., Gotoh M., Hirohashi S.;
 RT "Identification of human cadherin-14, a novel neurally specific type
 RT I cadherin, by protein interaction cloning.";
 RL J. Biol. Chem. 272:5226-5240(1997).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U59325; AAB02933.1; -;
 CC HSSP: P15116; INCI.
 CC PRINTS: PR00205; CADHERIN.

DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KW Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 24
 FT PROPEP 25 53
 FT CHAIN 54 790
 FT DOMAIN 54 608
 FT TRANSMEM 609 636
 FT DOMAIN 637 790
 FT REPEAT 54 159
 FT REPEAT 160 268
 FT REPEAT 269 383
 FT REPEAT 384 486
 FT REPEAT 487 608
 FT CARBOHYD 36 36
 FT CARBOHYD 255 255
 FT CARBOHYD 455 455
 FT CARBOHYD 536 536
 SO SEQUENCE 790 AA: 88072 MW: 5C7BDEB229B6EDCA CRC64;
 Query Match 57.1%; Score 60; DB 1; Length 790;
 Best Local Similarity 58.3%; Pred. No. 8,91e-01;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 DB 100 IFIIDTGTGDIH 111
 ID 3 IFVVDKNTGDIIN 14
 OY
 RESULT 14
 ID POLG_PYEVI STANDARD: PRT: 3027 AA.
 AC Q05057;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: 22.5 KD PROTEIN; 26 KD PROTEIN; 31 KD
 DE PROTEIN; PROBABLE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)].
 OS Parsnip yellow fleck virus (isolate P-121) (PYFV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Sequiviridae;
 OC Sequivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93107855.
 RA Turnbull-Ross A.D., Reavy B., Mayo M.A., Murrant A.F.;
 RT "The nucleotide sequence of parsnip yellow fleck virus: a plant
 RT picorna-like virus.";
 RL J. Gen. Virol. 73:3203-3211(1992).
 CC -1- SIMILARITY: SOME TO THE CMV AND TBV POLYPROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D14066; BAA03151.1; -;
 DR PIR: J01917; J01917.
 DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM: PF00910; RNA_helicase; 1.
 KW Polyprotein; ATP-binding; Coat protein; Transferase;
 KW RNA-directed RNA polymerase.
 FT NP_BIND 1467 1474
 FT VARIANT 962 962
 FT VARIANT 1373 1373
 SO SEQUENCE 3027 AA: 336242 MW: 0C41EB985F405BE2 CRC64;
 Query Match 55.2%; Score 58; DB 1; Length 3027;
 Best Local Similarity 40.0%; Pred. No. 2,21e+00;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 2246 FGVGVNDTIGIDV 2260

QY 1 FGIFVVDKNTGDIINI 15

RESULT 15
ID RNH2_HELPJ STANDARD: PRT: 209 AA.

AC 09JUR1: 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE RIBONUCLEASE HII (EC 3.1.26.4) (RNASE HII).
GN RNHB OR HP1323.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria: Proteobacteria: epsilon subdivision: Helicobacter group:
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99120557.
RA Alm R.A., Ling L.-S.L., Moll D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., Dejonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.W., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic-sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 367:176-180(1999).
CC -1- FUNCTION: Degrades the ribonucleotide moiety on RNA-DNA hybrid
CC -1- MOLECULES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: endonucleolytic cleavage to 5'-phospho-
CC -1- MONOESTER.
CC -1- COFACTOR: requires manganese for activity (by similarity).
CC -1- SUBCELLULAR LOCATION: cytoplasmic (potential).
CC -1- SIMILARITY: belongs to the RNase HII family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE001548; AAD06828.1; -
KW Hydrolyase; Nuclease; Endonuclease; Manganese.
SQ SEQUENCE 209 AA: 23078 MW: F06CDEAFC163D1D CRC64:

Query Match 54.3%; Score 57; DB 1; Length 209;
Best Local Similarity 46.2%; Pred. No. 3.44e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 67 GFVVKRSANEID 79
QY 2 GFVVDKNTGDIINI 14

Search completed: Sat May 13 07:00:04 2000
Job time : 109 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch: protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:00:24 2000; Mpsrch time 232.53 Seconds

Tabular output not generated. 4.473 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Perfect Score: 105

Sequence: 1 FGIFVVDKNTGDINI 15

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

sptrembl12
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phage 10:sp_plant 11:sp_ricent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.732; Variance 34.969; scale 0.764

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	90	85.7	993	11	035902	DESMOGLEIN 3 (FRAGMENT	4.79e-07
2	66	62.9	785	13	090763	CHICKEN CADHERIN-7.	1.24e-01
3	63	60.0	1329	5	076356	C45G7.6 PROTEIN.	5.11e-01
4	61	58.1	3503	5	024292	ADHERIN.	1.28e+00
5	60	57.1	214	2	055701	HYPOTHETICAL 24.5 KD P	2.02e+00
6	59	57.1	598	1	029907	PROBABLE TUNGSTEN-COAT	2.02e+00
7	59	56.2	154	1	058735	HYPOTHETICAL PROTEIN M	3.17e+00
8	59	56.2	245	1	058723	HYPOTHETICAL PROTEIN M	3.17e+00
9	59	56.2	624	10	048662	PV72.	3.17e+00
10	59	56.2	1033	2	033373	PILC PROTEIN.	3.17e+00
11	58	55.2	270	1	028514	CONSERVED HYPOTHETICAL	4.94e+00
12	58	55.2	555	10	0920W3	LACCASE (EC 1.10.3.2).	4.94e+00
13	58	55.2	593	10	065522	HYPOTHETICAL 68.6 KD P	4.94e+00
14	58	55.2	816	4	075284	KIAA0345-LIKE 9.	4.94e+00
15	58	55.2	936	4	0955H7	PROTODADHERIN ALPHA 5.	4.94e+00
16	58	55.2	930	13	095508	PARAXIAL PROTODADHERIN	4.94e+00
17	58	55.2	1035	13	057537	NE-PROTODADHERIN.	4.94e+00
18	58	55.2	1069	4	060245	PCDH7 (BH-PCDH)A.	4.94e+00
19	58	55.2	1072	4	060246	PCDH7 (BH-PCDH)B.	4.94e+00
20	58	55.2	1200	4	060247	PCDH7 (BH-PCDH)C.	4.94e+00

21	57	54.3	209	2	092JRI	RIBONUCLEASE HII.	7.66e+00
22	57	54.3	504	1	057984	504AA LONG HYPOTHETICA	7.66e+00
23	57	54.3	693	4	015066	OB-CADHERIN-2.	7.66e+00
24	57	54.3	792	13	093319	CADHERIN 11.	7.66e+00
25	57	54.3	794	13	093264	CADHERIN PRECURSOR.	7.66e+00
26	57	54.3	797	11	063418	PUTATIVE LIPOPROTEIN.	7.66e+00
27	57	54.3	798	2	050288	PUTATIVE LIPOPROTEIN.	7.66e+00
28	57	54.3	803	4	075283	KIAA0345-LIKE 8.	7.66e+00
29	57	54.3	814	4	075281	KIAA0345-LIKE 6.	7.66e+00
30	57	54.3	950	4	0915H6	PROTODADHERIN ALPHA 8.	7.66e+00
31	56	53.3	171	5	061171	CENTRIN 1 (FRAGMENT).	1.18e+01
32	56	53.3	205	5	027342	GLUTATHIONE PEROXIDASE	1.18e+01
33	56	53.3	423	10	065875	CELL DIVISION PROTEIN	1.18e+01
34	56	53.3	601	2	086691	PUTATIVE TRANSPORT SYS	1.18e+01
35	56	53.3	649	10	09XFN8	17F8.5	1.18e+01
36	56	53.3	788	4	09Y6N8	CADHERIN-10.	1.18e+01
37	56	53.3	831	5	P913B8	COSMID K1209.	1.18e+01
38	55	52.4	143	2	066117	HYPOTHETICAL 16.3 KD P	1.18e+01
39	55	52.4	261	13	042389	SUCRASE-ISOMALTASE (PR	1.18e+01
40	55	52.4	340	10	022706	F8A5.20 PROTEIN.	1.18e+01
41	55	52.4	569	14	069013	POLYMERASE (FRAGMENT).	1.18e+01
42	55	52.4	706	13	093869	GLYCOGEN SYNTHASE.	1.18e+01
43	55	52.4	790	13	097672	CHICKEN CADHERIN-6B.	1.18e+01
44	55	52.4	1196	2	045916	138RDA PROTEIN ASSOCIA	1.18e+01
45	55	52.4	1196	9	092X77	NTNH.	1.18e+01

ALIGNMENTS

RESULT	ID	035902	PREDIMINARY	PRT	993 AA.
AC	035902				
DT	01-JAN-1998	(TREMblrel. 05, Created)			
DT	01-JAN-1998	(TREMblrel. 05, Last sequence update)			
DT	01-NOV-1999	(TREMblrel. 12, Last annotation update)			
DE	DESMOGLEIN 3 (FRAGMENT).				
GN	DSG3.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
NC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
CC	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BALB/C;				
RA	ISHIKAWA H., LI K., UIRTO J.;				
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).				
DR	EMBL: U86016; AAB5091.1; -.				
DR	HSSP: P15116; INCU.				
DR	PROSITE: PS00232; CADHERIN; 2.				
DR	PFAM: PF00028; cadherin; 4.				
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.				
FT	NON-TER 993				
SO	SEQUENCE 993 AA; 10788 MW; 881794BD CRC32;				

Query Match	85.7%	Score 90; DB 11; Length 993;
Best Local Similarity	86.7%	Pred. No. 4.79e+07;
Matches 13; Conservative	0;	Mismatches 2; Indels 0; Gaps 0;

Db	97	FGIFVVDKNTGDINI 111
Qy	1	FGIFVVDKNTGDINI 15
RESULT	2	
ID	090763	PRELIMINARY; PRT; 785 AA.
AC	090763	
DT	01-NOV-1996	(TREMblrel. 01, Created)
DT	01-NOV-1996	(TREMblrel. 01, Last sequence update)
DE	CHICKEN CADHERIN-7.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;	
NC	Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.	

RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-WHITE LECHORN: TISSUE-BRAIN:
 RX MEDLINE: 95309115.
 RA NAKAGAWA S., TAKEICHI M.:
 RT "Neural crest cell-cell adhesion controlled by sequential and
 subpopulation-specific expression of novel cadherins."
 RL Development 121:1321-1332(1995).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: D42150; BAA07721.1; -.
 DR HSSP: P15116; INCD.
 DR PROSITE: PS00232; CADHERIN: 3.
 DR PFAM: PF00028; cadherin: 5.
 DR PFAM: PF01049; Cadherin_C-term: 1.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SO SEQUENCE 785 AA; 87177 MW; F27A6881 CRC32;

Query Match 62.9%; Score 66; DB 13; Length 785;
 Best Local Similarity 61.5%; Pred. No. 1.24e+01;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 SIFIDNTGDIH 105
 Qy 2 GIFVVDKNTGDI 14

RESULT 3 PRELIMINARY; PRT: 1329 AA.
 AC 076356;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE C4567.6 PROTEIN.
 GN C4567.6
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 BONTLED J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
 JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RT Nature 368:32-38(1994).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DANTE M., WAMBLEY P.;
 RT "The sequence of C. elegans cosmid C4567."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF067611; AAC19184.1; -.
 DR PROSITE: PS00232; CADHERIN: 1.
 DR PFAM: PF00028; cadherin: 3.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SO SEQUENCE 1329 AA; 146518 MW; B161D39E CRC32;

Query Match 60.0%; Score 63; DB 5; Length 1329;
 Best Local Similarity 72.7%; Pred. No. 5.11e+01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1288 IFVVDKNTGDI 1298
 Qy 3 IFVVDKNTGDI 13

RESULT 4 PRELIMINARY; PRT: 3503 AA.
 AC 024292;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ADHERIN.
 GN ADHERIN.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95324813.
 RA CLARK H.F., BRENTNUP D., SCHNEITZ K., BIBER A., GOODMAN C., NOLL M.;
 RT "Dachsous encodes a member of the cadherin superfamily that controls
 imaginal disc morphogenesis in Drosophila."
 RL Genes Dev. 9:1530-1542(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA NOLL M.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: L08811; AAA79329.2; -.
 DR HSSP: P15116; INCD.
 DR FLYBASE: FBgn0000497; ds.
 DR PROSITE: PS00232; CADHERIN: 20.
 DR PFAM: PF00028; cadherin: 25.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SO SEQUENCE 3503 AA; 379719 MW; 16030CA5 CRC32;

Query Match 58.1%; Score 61; DB 5; Length 3503;
 Best Local Similarity 58.1%; Pred. No. 1.28e+00;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 713 GIFRIDRSTGEI 724
 Qy 2 GIFVVDKNTGDI 13

RESULT 5 PRELIMINARY; PRT: 214 AA.
 AC 055701;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE HYPOTHETICAL 24.5 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 96127529.
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
 RA SUGIURA M., TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb


```

NN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GC-653.
RX MEDLINE: 98129088.
RA BAECKMAN M., KÄLLSTRÖM H., JONSSON A.B.;
RT "The phase-variable plus-associated protein PilC is commonly
RT expressed in clinical isolates of Neisseria gonorrhoeae, and shows
RT sequence variability among strains."
RL Microbiology 144:149-156(1998).
DR EMBL: AJ001121; CA045457.1; -.
SO SEQUENCE 1033 AA; 112936 MW; B0593053 CRC32;

Dd Query Match 56.2%; Score 59; DB 2; Length 1033;
Best Local Similarity 46.7%; Pred. No. 3,17e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0

Oy 1 GFIEFVDKNTGDINI 15
::: ||| ||| ::

RESULT 11
ID -028514 PRELIMINARY; PRT; 270 AA.
AC -028514;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF1760.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.T., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBEK R., GOCAVNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYRES S.M.,
RA SAOM P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MACON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL: AE000981; AAB89491.1; -.
DR HSSP: O57997; LMJH.
DR TIGR: AF1760; -.
DR PFAM: PF00582; Usp_1.
DR Hypothetical protein.
KW Hypothetical protein.
SO SEQUENCE 270 AA; 29554 MW; E5FA8A9B CRC32;

Dd Query Match 55.2%; Score 58; DB 1; Length 270;
Best Local Similarity 40.0%; Pred. No. 4.94e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 165 YAKFVVKKTGGELHI 179
:: ||| | ||| |
Oy 1 GFIEFVDKNTGDINI 15

RESULT 12
ID 0920W3 PRELIMINARY; PRT; 555 AA.
AC 0920W3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TACCAF (PFC 1.10.3.2).

```

GN LAC3.
 OS Populus trichocarpa (Western balsam poplar).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC Euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: rosidae: eurosids I: Malpighiales: Salicaceae: Populus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. TRICHOBEL; TISSUE=XYLEM;
 RX MEDLINE: 99115479.
 RA RANOCIA P., MCDUGALL G., HAWKINS S., STERJADES R., BORDERIES G.,
 RA STEWART D., CABANES-MACHEAU M., BOUDER A.M., GOFENR D.,
 RT "Biochemical characterization, molecular cloning and expression of
 RT laccases - a divergent gene family - in poplar."
 RL EUR. J. Biochem. 259:485-495(1999).
 DR EMBL: Y13771; CAAT74103.1; -
 DR HSSP: P37064; JASP.
 DR MENDEL: 34836; Popbe;1457;34836.
 KM OXIDOREDUCTASE.
 SQ SEQUENCE 555 AA: 60890 MW: 982560EC CRC32:
 Query Match 55.2%; Score 58; DB 10; Length 555;
 Best Local Similarity 42.9%; Pred. No. 4.94e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 463 ENFVVGKIGNFD 476
 1 FGVVDKNTGDIIN 14
 QY 1 FGVVDKNTGDIIN 14
 RESULT 13
 ID 065522 PRELIMINARY: PRT: 593 AA.
 AC 065522;
 DT 01-AUG-1998 (TEMBREL. 07, Created)
 DT 01-AUG-1998 (TEMBREL. 07, Last sequence update)
 DT 01-NOV-1999 (TEMBREL. 12, Last annotation update)
 DE HYPOHETICAL 68.6 KD PROTEIN.
 GN PAD1.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC Euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: rosidae: eurosids II: Brassicales: Brassicaceae:
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., BENES V., RECHMANN S., BORKOVA D., ANSORGE W., HOEISEL J.,
 RA MEWES H.W., MAYER K.F.X., SCHUELLER C.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL022537; CA118582.1; -
 DR MENDEL: 29102; Arch;3426;29102.
 KW Hypochemical protein.
 SU SEQUENCE 593 AA: 68632 MW: 9EA10B88 CRC32:
 Query Match 55.2%; Score 58; DB 10; Length 593;
 Best Local Similarity 35.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
 Db 528 FAVEVLEKIDPLDK 541
 1 FGVVDKNTGDIIN 14
 QY 1 FGVVDKNTGDIIN 14
 RESULT 14
 ID 075284 PRELIMINARY: PRT: 816 AA.
 AC 075284;
 DT 01-NOV-1998 (TEMBREL. 08, Created)
 DT 01-NOV-1998 (TEMBREL. 08, Last sequence update)
 DT 01-NOV-1999 (TEMBREL. 12, Last annotation update)
 DE KIA0345-LIKE 9.
 GN PCDH-ALPHA5.

OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homiinae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KIMBERLY W., BONDOR M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
 RA KADNER K., MIGUEL T., MILLER C., PILLUCK S., POLLARD M., ROJESKI H.,
 RA SUBRAMANIAN S., MARTIN C.H.;
 RT "Sequencing of human chromosome 5."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA RICE D.O.;
 RT "Large Scale Sequence Analysis and Annotation with the Sequence
 RT Comparison Analysis (SCAN) System."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99308636.
 RA WU O., MANIATIS T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AC005609; AAC34321.1; -
 DR EMBL: AF152483; AAD3744.1; -
 DR PROSITE: PS00232; CADHERIN; 5.
 DR PFAM: PF00028; cadherin; 5.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 816 AA: 89348 MW: 3F41D941 CRC32:
 Query Match 55.2%; Score 58; DB 4; Length 816;
 Best Local Similarity 41.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 Db 292 FIINSNTGEIKV 303
 4 FVVDKNTGDIIN 15
 QY 4 FVVDKNTGDIIN 15
 RESULT 15
 ID 0955H7 PRELIMINARY: PRT: 936 AA.
 AC 0955H7;
 DT 01-NOV-1999 (TEMBREL. 12, Created)
 DT 01-NOV-1999 (TEMBREL. 12, Last sequence update)
 DT 01-NOV-1999 (TEMBREL. 12, Last annotation update)
 DE PROTOCADHERIN ALPHA 5.
 GN PCDH-ALPHA5.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
 OC Eutheria: Primates: Catarrhini: Homiinae: Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 99308636.
 RA WU O., MANIATIS T.;
 RT "A striking organization of a large family of human neural cadherin-
 RT like cell adhesion genes."
 RL Cell 97:779-790(1999).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF152313; AAD43707.1; -
 DR PROSITE: PS00232; CADHERIN; 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 936 AA: 102048 MW: 387ECF8D CRC32:
 Query Match 55.2%; Score 58; DB 4; Length 936;
 Best Local Similarity 41.7%; Pred. No. 4.94e+00;
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 Db 292 FIINSNTGEIKV 303
 4 FVVDKNTGDIIN 15
 QY 4 FVVDKNTGDIIN 15

Mon May 15 08:08:11 2000

US-08-991-628-2a.rspt

Page 6

Search completed: Sat May 13 07:04:30 2000
Job time : 246 secs.

[W] [E] [S] [E] [H]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 06:06:16 2000: Maspar time 52.14 Seconds

Tabular output not generated. 3.727 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Perfect Score: 105

Sequence: 1 FGIFFVDKNTGDIINI 15

Scoring table: PAM 150
Gap 15

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-issued

1:5A.COMB 2:5B.COMB 3:PCT.COMB 4:backfiles1

Statistics: Mean 17.421; Variance 58.153; scale 0.300

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	105	100.0	15	2	US-08-400-	Sequence 2, Applicatio	2.01e-04
2	57	54.3	615	2	US-08-738-	Sequence 12, Applicati	3.04e+01
3	57	54.3	693	2	US-08-738-	Sequence 6, Applicatio	3.04e+01
4	57	54.3	796	2	US-08-738-	Sequence 2, Applicatio	3.04e+01
5	57	54.3	796	2	US-08-188-	Sequence 36, Applicatio	3.04e+01
6	57	54.3	796	2	US-08-738-	Sequence 4, Applicatio	3.04e+01
7	57	54.3	796	1	US-08-332-	Sequence 52, Applicati	3.04e+01
8	57	54.3	796	1	US-08-332-	Sequence 58, Applicati	3.04e+01
9	57	54.3	797	1	US-08-268-	Sequence 112, Applicat	3.04e+01
10	57	54.3	797	2	US-08-453-	Sequence 112, Applicat	3.04e+01
11	57	54.3	797	2	US-08-453-	Sequence 112, Applicat	3.04e+01
12	57	54.3	797	3	PCT-US95-0	Sequence 112, Applicat	3.04e+01
13	57	54.3	837	2	US-08-472-	Sequence 6, Applicatio	3.04e+01
14	57	54.3	837	2	US-08-472-	Sequence 7, Applicatio	3.04e+01
15	57	54.3	837	2	US-08-474-	Sequence 7, Applicatio	3.04e+01
16	56	53.3	433	2	US-08-883-	Sequence 2, Applicatio	3.82e+01
17	56	53.3	511	2	US-08-537-	Sequence 4, Applicatio	3.82e+01
18	56	53.3	511	2	US-08-537-	Sequence 34, Applicati	3.82e+01
19	56	53.3	515	2	US-08-537-	Sequence 3, Applicatio	3.82e+01
20	56	53.3	713	1	US-08-332-	Sequence 56, Applicati	3.82e+01
21	56	53.3	713	1	US-08-332-	Sequence 62, Applicati	3.82e+01
22	56	53.3	713	1	US-08-188-	Sequence 62, Applicati	3.82e+01
23	55	52.4	878	1	US-08-237-	Sequence 2, Applicatio	4.81e-01

ALIGNMENTS

24	55	52.4	878	3	PCT-US95-0	Sequence 2, Applicatio	4.81e-01
25	55	52.4 <td>884<td>2<td>US-08-474-</td><td>Sequence 8, Applicatio<th>4.81e-01</th></td></td></td>	884 <td>2<td>US-08-474-</td><td>Sequence 8, Applicatio<th>4.81e-01</th></td></td>	2 <td>US-08-474-</td> <td>Sequence 8, Applicatio<th>4.81e-01</th></td>	US-08-474-	Sequence 8, Applicatio <th>4.81e-01</th>	4.81e-01
26	55	52.4 <td>884<td>2<td>US-08-474-</td><td>Sequence 8, Applicatio<th>4.81e-01</th></td></td></td>	884 <td>2<td>US-08-474-</td><td>Sequence 8, Applicatio<th>4.81e-01</th></td></td>	2 <td>US-08-474-</td> <td>Sequence 8, Applicatio<th>4.81e-01</th></td>	US-08-474-	Sequence 8, Applicatio <th>4.81e-01</th>	4.81e-01
27	55	52.4 <td>884<td>2<td>US-08-472-</td><td>Sequence 7, Applicatio<th>4.81e-01</th></td></td></td>	884 <td>2<td>US-08-472-</td><td>Sequence 7, Applicatio<th>4.81e-01</th></td></td>	2 <td>US-08-472-</td> <td>Sequence 7, Applicatio<th>4.81e-01</th></td>	US-08-472-	Sequence 7, Applicatio <th>4.81e-01</th>	4.81e-01
28	54	51.4 <td>246<td>2<td>US-08-415-</td><td>Sequence 31, Applicati<th>6.04e-01</th></td></td></td>	246 <td>2<td>US-08-415-</td><td>Sequence 31, Applicati<th>6.04e-01</th></td></td>	2 <td>US-08-415-</td> <td>Sequence 31, Applicati<th>6.04e-01</th></td>	US-08-415-	Sequence 31, Applicati <th>6.04e-01</th>	6.04e-01
29	53	50.5 <td>199<td>2<td>US-08-211-</td><td>Sequence 6, Applicatio<th>7.57e-01</th></td></td></td>	199 <td>2<td>US-08-211-</td><td>Sequence 6, Applicatio<th>7.57e-01</th></td></td>	2 <td>US-08-211-</td> <td>Sequence 6, Applicatio<th>7.57e-01</th></td>	US-08-211-	Sequence 6, Applicatio <th>7.57e-01</th>	7.57e-01
30	53	50.5 <td>794<td>1<td>US-08-332-</td><td>Sequence 54, Applicati<th>7.57e-01</th></td></td></td>	794 <td>1<td>US-08-332-</td><td>Sequence 54, Applicati<th>7.57e-01</th></td></td>	1 <td>US-08-332-</td> <td>Sequence 54, Applicati<th>7.57e-01</th></td>	US-08-332-	Sequence 54, Applicati <th>7.57e-01</th>	7.57e-01
31	53	50.5 <td>794<td>1<td>US-08-332-</td><td>Sequence 54, Applicati<th>7.57e-01</th></td></td></td>	794 <td>1<td>US-08-332-</td><td>Sequence 54, Applicati<th>7.57e-01</th></td></td>	1 <td>US-08-332-</td> <td>Sequence 54, Applicati<th>7.57e-01</th></td>	US-08-332-	Sequence 54, Applicati <th>7.57e-01</th>	7.57e-01
32	53	50.5 <td>794<td>1<td>US-08-188-</td><td>Sequence 60, Applicati<th>7.57e-01</th></td></td></td>	794 <td>1<td>US-08-188-</td><td>Sequence 60, Applicati<th>7.57e-01</th></td></td>	1 <td>US-08-188-</td> <td>Sequence 60, Applicati<th>7.57e-01</th></td>	US-08-188-	Sequence 60, Applicati <th>7.57e-01</th>	7.57e-01
33	53	50.5 <td>2647<td>2<td>US-08-583-</td><td>Sequence 8, Applicatio<th>7.57e-01</th></td></td></td>	2647 <td>2<td>US-08-583-</td><td>Sequence 8, Applicatio<th>7.57e-01</th></td></td>	2 <td>US-08-583-</td> <td>Sequence 8, Applicatio<th>7.57e-01</th></td>	US-08-583-	Sequence 8, Applicatio <th>7.57e-01</th>	7.57e-01
34	53	50.5 <td>2647<td>2<td>US-08-779-</td><td>Sequence 8, Applicatio<th>7.57e-01</th></td></td></td>	2647 <td>2<td>US-08-779-</td><td>Sequence 8, Applicatio<th>7.57e-01</th></td></td>	2 <td>US-08-779-</td> <td>Sequence 8, Applicatio<th>7.57e-01</th></td>	US-08-779-	Sequence 8, Applicatio <th>7.57e-01</th>	7.57e-01
35	52	49.5 <td>535<td>2<td>US-08-007-</td><td>Sequence 2, Applicatio<th>9.48e-01</th></td></td></td>	535 <td>2<td>US-08-007-</td><td>Sequence 2, Applicatio<th>9.48e-01</th></td></td>	2 <td>US-08-007-</td> <td>Sequence 2, Applicatio<th>9.48e-01</th></td>	US-08-007-	Sequence 2, Applicatio <th>9.48e-01</th>	9.48e-01
36	52	49.5 <td>1026<td>1<td>US-08-268-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>1<td>US-08-268-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	1 <td>US-08-268-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-08-268-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
37	52	49.5 <td>1026<td>1<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>1<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	1 <td>US-08-453-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-08-453-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
38	52	49.5 <td>1026<td>2<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>2<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	2 <td>US-08-453-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-08-453-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
39	52	49.5 <td>1026<td>1<td>US-07-998-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>1<td>US-07-998-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	1 <td>US-07-998-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-07-998-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
40	52	49.5 <td>1026<td>1<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>1<td>US-08-453-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	1 <td>US-08-453-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-08-453-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
41	52	49.5 <td>1026<td>3<td>PCT-US93-1</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1026 <td>3<td>PCT-US93-1</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	3 <td>PCT-US93-1</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	PCT-US93-1	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
42	52	49.5 <td>1203<td>1<td>US-07-998-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td></td>	1203 <td>1<td>US-07-998-</td><td>Sequence 95, Applicati<th>9.48e-01</th></td></td>	1 <td>US-07-998-</td> <td>Sequence 95, Applicati<th>9.48e-01</th></td>	US-07-998-	Sequence 95, Applicati <th>9.48e-01</th>	9.48e-01
43	52	49.5 <td>1203<td>1<td>US-08-268-</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td></td>	1203 <td>1<td>US-08-268-</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td>	1 <td>US-08-268-</td> <td>Sequence 103, Applicat<th>9.48e-01</th></td>	US-08-268-	Sequence 103, Applicat <th>9.48e-01</th>	9.48e-01
44	52	49.5 <td>1203<td>3<td>PCT-US93-1</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td></td>	1203 <td>3<td>PCT-US93-1</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td>	3 <td>PCT-US93-1</td> <td>Sequence 103, Applicat<th>9.48e-01</th></td>	PCT-US93-1	Sequence 103, Applicat <th>9.48e-01</th>	9.48e-01
45	52	49.5 <td>1203<td>1<td>US-08-453-</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td></td>	1203 <td>1<td>US-08-453-</td><td>Sequence 103, Applicat<th>9.48e-01</th></td></td>	1 <td>US-08-453-</td> <td>Sequence 103, Applicat<th>9.48e-01</th></td>	US-08-453-	Sequence 103, Applicat <th>9.48e-01</th>	9.48e-01

RESULT 1
ID US-08-400-796-2 STANDARD: PRT: 15 AA.

Sequence 2, Application US/08400796

Sequence 2, Application US/08400796
Patent No. 5874531

GENERAL INFORMATION:
APPLICANT: STROMINGER, JACK L.
TITLE OF INVENTION: IDENTIFICATION OF SELF AND NON-SELF
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/400,796
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: TWOMEY, MICHAEL J.
REGISTRATION NUMBER: 38,349
REFERENCE/DOCKET NUMBER: H0498/7015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO

ID	RESULT	4			PRT:	796 AA.
XX	US-08-738-349-2		STANDARD:			
XX	xxxxxx					
XX						
DE	Sequence 2, Application US/08738349					
CC	Sequence 2, Application US/08738349					
CC	Patent No. 5869638					
CC	GENERAL INFORMATION:					
CC	APPLICANT: Takeshita, Sunao					
CC	APPLICANT: Okazaki, Makoto					
CC	APPLICANT: Kawai, Shinji					
CC	APPLICANT: Tsujimura, Atsushi					
CC	APPLICANT: Amann, Egon					
CC	TITLE OF INVENTION: Bone-Related Cadherin-Like Protein and					
CC	TITLE OF INVENTION: Process for Its Production					
CC	NUMBER OF SEQUENCES: 12					
CC	CORRESPONDENCE ADDRESS:					
CC	ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &					
CC	ADDRESSEE: Dunner					
CC	STREET: 1300 I Street, N.W.					
CC	CITY: Washington					
CC	STATE: D.C.					
CC	COUNTRY: USA					
CC	ZIP: 20005-3115					
CC	COMPUTER READABLE FORM:					
CC	MEDIUM TYPE: Floppy disk					
CC	COMPUTER: IBM PC compatible					
CC	OPERATING SYSTEM: PC-DOS/MS-DOS					
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25					
CC	CURRENT APPLICATION DATA:					
CC	APPLICATION NUMBER: US/08/738,349					
CC	FILING DATE: 25-OCT-1996					
CC	CLASSIFICATION: 435					
CC	PRIOR APPLICATION DATA:					
CC	APPLICATION NUMBER: US 08/364,439					
CC	FILING DATE:					
CC	APPLICATION NUMBER: US 08/112,061					
CC	FILING DATE: 26-AUG-1993					
CC	ATTORNEY/AGENT INFORMATION:					
CC	NAME: Barker, M. P.					
CC	REGISTRATION NUMBER: 32,013					
CC	REFERENCE/DOCKET NUMBER: 02481.1323-00000					
CC	TELECOMMUNICATION INFORMATION:					
CC	TELEPHONE: 202-408-4000					
CC	TELEFAX: 202-408-4400					
CC	INFORMATION FOR SEQ ID NO: 2:					
CC	SEQUENCE CHARACTERISTICS:					
CC	LENGTH: 796 amino acids					
CC	TYPE: amino acid					
CC	TOPOLOGY: linear					
CC	MOLECULE TYPE: protein					
SQ	SEQUENCE 796 AA; 88112 MM; 3181201 CN;					
DB	100 IFVIDKSGNTH 111					
OY	3 IFVVDKNTGDI 14					
XX	Query Match	54.3%	Score 57;	DB 2;	Length 796;	
XX	Best Local Similarity	50.0%;	Pred. No. 3.04e+01;			
XX	Matches	6;	Conservative	5;	Mismatches	1;
XX					Indels	0;
XX					Gaps	0;
XX	RESULT	5				
XX	ID	US-08-188-228-58	STANDARD:	PRT:	796 AA.	
XX	AC	xxxxxx				
XX	DT					
XX	TT					
XX						

```

DE      Sequence 58, Application US/08188228
XX
CC      Sequence 59, Application US/08188228
CC      Patent No. 5597725
CC      GENERAL INFORMATION:
CC      APPLICANT: Suzuki, Shintaro
CC      TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
CC      NUMBER OF SEQUENCES: 62
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
CC      ADDRESSEE: Borun
CC      CITY: Chicago
CC      STATE: Illinois
CC      COUNTRY: USA
CC      ZIP: 60606
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patent Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/188,228
CC      FILING DATE:
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/049,460
CC      FILING DATE: 19 APR 1993
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 07/872,643
CC      FILING DATE: 17 APR 1992
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: No. 5597725and, Greta E.
CC      REGISTRATION NUMBER: 35,302
CC      REFERENCE/DOCKET NUMBER: 31340
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (312) 474-6300
CC      TELEFAX: (312) 474-0448
CC      TELEX: 25-3856
CC      INFORMATION FOR SEQ ID NO: 58:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 796 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      SEQUENCE 796 AA; 88049 MW; 3168744 CN;
SQ      QUERY MATCH
          54.3%; Score 57; DB 1; Length 796;
          Best Local Similarity 50.0%; Pct. No. 3.04e+01;
          Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0
DB      100 IFVIDKSGNTH 111
          |||:|::|:|
QY      3 IFVVDKNTGTDIN 14
RESULT 6
ID      US-08-738-349-4      STANDARD:      PRT:      796 AA.
XX      AC      xxxxxx
XX      DT
XX      DE      Sequence 4, Application US/08738349
XX      CC      Sequence 4, Application US/08738349
XX      CC      Patent No. 5869638
XX      CC      GENERAL INFORMATION:
XX      CC      APPLICANT: Takeshita, Sunao
XX      CC      APPLICANT: Okazaki, Makoto
XX      CC      APPLICANT: Kawai, Shintji
XX      CC      APPLICANT: Tsujimura, Atsushi
XX      CC      APPLICANT: Amann, Egon

```


XX DE Sequence 112, Application US/08453702A
 CC XX Sequence 112, Application US/08453702A
 CC Patent No. 5891706
 CC GENERAL INFORMATION:
 CC APPLICANT: Suzuki, Shintaro
 CC TITLE OF INVENTION: Protocadherin Materials and Methods
 CC NUMBER OF SEQUENCES: 115
 CC CORRESPONDENCE ADDRESS: 115
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
 CC STREET: 233 South Wacker, 6300 Sears Tower
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60606
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/453,702A
 CC FILING DATE:
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: No. 5891706and, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 312/474-6300
 CC TELEFAX: 312/474-0448
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 112:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 797 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 797 AA: 87474 MW: 3358246 CN:
 SO
 Query Match 54.3%; Score 57; DB 2; Length 797;
 Best Local Similarity 46.7%; Pred. No. 3.04e+01;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 288 FGFVVDKNTGDI 302
 QY 1 FGFVVDKNTGDI 15
 RESULT 12 STANDARD: PRT: 797 AA.
 ID PCV-0595-08071-112
 AC xxxxxx
 DE Sequence 112, Application PC/TUS9508071
 CC GENERAL INFORMATION:
 CC APPLICANT: Suzuki, Shintaro
 CC TITLE OF INVENTION: Protocadherin Materials and Methods
 CC NUMBER OF SEQUENCES: 115
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
 CC STREET: 6300 Sears Tower, 233 S. Wacker Drive
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60606

CC CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US95/08071
 CC FILING DATE:
 CC CLASSIFICATION:
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US93/12588
 CC FILING DATE: 23 DEC 1993
 CC PRIOR APPLICATION DATA: US 07/998,003
 CC APPLICATION NUMBER: 29 DEC 1992
 CC FILING DATE: 29 DEC 1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Noland, Greta E.
 CC REGISTRATION NUMBER: 35,302
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 312/474-6300
 CC TELEFAX: 312/474-0448
 CC TELEX: 25-3856
 CC INFORMATION FOR SEQ ID NO: 112:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 797 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: Protein
 CC SEQUENCE 797 AA: 87474 MW: 3358246 CN:
 SO
 Query Match 54.3%; Score 57; DB 3; Length 797;
 Best Local Similarity 46.7%; Pred. No. 3.04e+01;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 288 FGFVVDKNTGDI 302
 QY 1 FGFVVDKNTGDI 15
 RESULT 13 STANDARD: PRT: 837 AA.
 ID US-08-472-481-6
 AC xxxxxx
 DE Sequence 6, Application US/08472481
 CC GENERAL INFORMATION:
 CC APPLICANT: Ranscht, Barbara
 CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
 CC NUMBER OF SEQUENCES: 8
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Campbell and Flores
 CC STREET: 4370 La Jolla Village Drive, Suite 700
 CC CITY: San Diego
 CC STATE: California
 CC COUNTRY: United States
 CC ZIP: 92122
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/472,481
 CC FILING DATE: 07-JUN-1995
 CC CLASSIFICATION: 424
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/213,361

CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1686
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DE Sequence 7, Application US/08474067
|||:::|
ID US-08-474-067-7 STANDARD; PRT: 837 AA.
XX xxxxxx
AC
DT
DT
DE

Sequence 7, Application US/08474067
Patent No. 581518
GENERAL INFORMATION:
CC APPLICANT: Ranscht, Barbara
CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell and Flores
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: United States
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/474,067
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/213,361
CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1682
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DE Sequence 7, Application US/08474068A
|||:::|
ID US-08-474-068A-7 STANDARD; PRT: 837 AA.
XX xxxxxx
AC
DT
DT
DE

Sequence 7, Application US/08474068A
Patent No. 5837525
GENERAL INFORMATION:
CC APPLICANT: Ranscht, Barbara
CC TITLE OF INVENTION: T-Cadherin Adhesion Molecule
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell & Flores LLP
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: United States
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/474,068A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/213,361
CC FILING DATE: 14-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/607,293
CC FILING DATE: 30-OCT-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 1683
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 837 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 837 AA: 92281 MW: 3540645 CN:

Db 159 GIFFIERETG 168
|||:::|
OY 2 GIFFVDKNTG 11

Query Match 54.3%; Score 57; DB 2; Length 837;
Best Local Similarity 50.0%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Mon May 15 08:08:03 2000

US-08-991-628-2a.ra1

Page 8

Search completed: Sat May 13 06:07:19 2000
Job time : 63 secs.

THIS PAGE BLANK (USPTO)

 MWSELEF (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:10:47 2000; Maspar time 3.01 Seconds
 Tabular output not generated. 117,892 Million cell updates/sec

Title: >US-08-991-628-3
 Description: (1-15) from US08991628.pep
 Perfect Score: 96
 Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table:
 PAM 150
 Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseq35
 1:geneseqp

Statistics: Mean 18.170; Variance 51.971; scale 0.350

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	96	100.0	15	W04843	Self epitope of desmog	3.93e-04
2	96	100.0	15	W64815	Desmoglein-3 190-204	3.93e-04
3	96	100.0	15	W78814	Desmoglein 3 protein f	3.93e-04
4	96	100.0	15	W07908	Pemphigus vulgaris ant	3.93e-04
5	96	100.0	15	R30742	Human pemphigus vulgar	3.93e-04
6	87	90.6	778	W15469	Pemphigus foliaceus an	5.50e-03
7	77	80.2	263	W13010	Segment of desmosomal	9.68e-02
8	77	80.2	560	W13009	Segment of desmosomal	9.68e-02
9	77	80.2	560	W13009	Segment of desmosomal	9.68e-02
10	54	56.3	878	R85487	Human E-cadherin precu	4.96e+01
11	54	56.3	2233	W48711	HPV-3 JS Isolate Wild	4.96e+01
12	54	56.3	2233	W48713	HPV-3 Vero cp45 vacci	4.96e+01
13	54	56.3	2233	W48712	HPV-3 F81 cp45 vacci	4.96e+01
14	53	55.2	456	W82841	Human cerebral protein	6.41e+01
15	49	51.0	572	W40843	Bilirubin oxidase	1.75e+02
16	49	51.0	916	W13129	Full length human cadh	1.75e+02
17	49	51.0	916	W25658	Human cadherin-4	1.75e+02
18	49	51.0	2408	R24307	Translation of ORF 3 c	1.75e+02
19	48	50.0	26	W60192	Bacteriophage spo1 Pol	2.25e+02
20	48	50.0	528	W20908	H. pylori inner membra	2.25e+02
21	48	50.0	2016	W23994	Human hml sodium chann	2.25e+02
22	48	50.0	2019	R67913	Cardiac sodium channel	2.25e+02
23	48	50.0	2020	R06584	Cardiac sodium channel	2.25e+02

24	47	49.0	198	1	R06643	Mamalian growth hormon	2.87e-02
25	47	49.0 <td>308</td> <td>1</td> <td>W75006</td> <td>Human secreted protein</td> <td>2.87e-02</td>	308	1	W75006	Human secreted protein	2.87e-02
26	47	49.0 <td>330</td> <td>1</td> <td>W89745</td> <td>Staphylococcus aureus</td> <td>2.87e-02</td>	330	1	W89745	Staphylococcus aureus	2.87e-02
27	47	49.0 <td>399</td> <td>1</td> <td>W74883</td> <td>Human secreted protein</td> <td>2.87e-02</td>	399	1	W74883	Human secreted protein	2.87e-02
28	47	49.0 <td>573</td> <td>1</td> <td>R75741</td> <td>B31 outer surface prot</td> <td>2.87e-02</td>	573	1	R75741	B31 outer surface prot	2.87e-02
29	47	49.0 <td>700</td> <td>1</td> <td>R75731</td> <td>B. burgdorferi strain</td> <td>2.87e-02</td>	700	1	R75731	B. burgdorferi strain	2.87e-02
30	47	49.0 <td>716</td> <td>1</td> <td>R30730</td> <td>B. burgdorferi 79 kd a</td> <td>2.87e-02</td>	716	1	R30730	B. burgdorferi 79 kd a	2.87e-02
31	47	49.0 <td>756</td> <td>1</td> <td>R30177</td> <td>ppl.</td> <td>2.87e-02</td>	756	1	R30177	ppl.	2.87e-02
32	47	49.0 <td>4472</td> <td>1</td> <td>R97246</td> <td>Virulence gene cluster</td> <td>2.87e-02</td>	4472	1	R97246	Virulence gene cluster	2.87e-02
33	46	47.9 <td>33</td> <td>1</td> <td>W79334</td> <td>Staphylococcus aureus</td> <td>3.66e-02</td>	33	1	W79334	Staphylococcus aureus	3.66e-02
34	46	47.9 <td>298</td> <td>1</td> <td>R97629</td> <td>Human SLAMF2 T-cell co</td> <td>3.66e-02</td>	298	1	R97629	Human SLAMF2 T-cell co	3.66e-02
35	46	47.9 <td>305</td> <td>1</td> <td>R97630</td> <td>Human SLAMF3 T-cell co</td> <td>3.66e-02</td>	305	1	R97630	Human SLAMF3 T-cell co	3.66e-02
36	46	47.9 <td>307</td> <td>1</td> <td>R97631</td> <td>Human SLAMF4 T-cell co</td> <td>3.66e-02</td>	307	1	R97631	Human SLAMF4 T-cell co	3.66e-02
37	46	47.9 <td>335</td> <td>1</td> <td>R97638</td> <td>Human SLAMF1 T-cell co</td> <td>3.66e-02</td>	335	1	R97638	Human SLAMF1 T-cell co	3.66e-02
38	46	47.9 <td>369</td> <td>1</td> <td>W07080</td> <td>Human cancer associate</td> <td>3.66e-02</td>	369	1	W07080	Human cancer associate	3.66e-02
39	46	47.9 <td>459</td> <td>1</td> <td>R13458</td> <td>P. denitrificans COB G.</td> <td>3.66e-02</td>	459	1	R13458	P. denitrificans COB G.	3.66e-02
40	46	47.9 <td>481</td> <td>1</td> <td>W34554</td> <td>MYTL glycosidase 29G.</td> <td>3.66e-02</td>	481	1	W34554	MYTL glycosidase 29G.	3.66e-02
41	46	47.9 <td>486</td> <td>1</td> <td>W38455</td> <td>Yeast RNA-binding prot</td> <td>3.66e-02</td>	486	1	W38455	Yeast RNA-binding prot	3.66e-02
42	46	47.9 <td>779</td> <td>1</td> <td>R98226</td> <td>Rat neuronal protein k</td> <td>3.66e-02</td>	779	1	R98226	Rat neuronal protein k	3.66e-02
43	46	47.9 <td>821</td> <td>1</td> <td>R99579</td> <td>Calpain large subunit</td> <td>3.66e-02</td>	821	1	R99579	Calpain large subunit	3.66e-02
44	46	47.9 <td>1275</td> <td>1</td> <td>R13516</td> <td>P. denitrificans COB N.</td> <td>3.66e-02</td>	1275	1	R13516	P. denitrificans COB N.	3.66e-02
45	46	47.9 <td>1276</td> <td>1</td> <td>R35199</td> <td>Mouse multidrug resist</td> <td>3.66e-02</td>	1276	1	R35199	Mouse multidrug resist	3.66e-02

ALIGNMENTS

RESULT 1
 ID W04843 standard: peptide: 15 AA.
 AC W04843;
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KM Tolerisation: self-epitope; antigen; autoimmune disease.
 KM autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KM pemphigus vulgaris; desmoglein; multiple sclerosis;
 KM herpes simplex virus; adenovirus; phosphoanionomutase;
 KM human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KM influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PF 07-MAR-1996; U03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Mucherferfennly KW;
 DR WPI: 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 39; 58pp; English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 190-204)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SO Sequence 15 AA:

Query Match 100.0%; Score 96; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.93e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 LNSKIAFKIVSOEPA 15
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 2
ID W64815 standard: peptide; 15 AA.
AC W64815:
DT 29-SEP-1998 (first entry)
DE Desmoglein-3 190-204.
KW Desmoglein: DG: gene therapy; pemphigus vulgaris; microparticle:
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PD 21-JUL-1998.
PF 22-JAN-1997; 787547.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANCAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS;
DR WPI: 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PS expression of proteins e.g. in gene therapy
PS Disclosure: Column 4; 42pp: English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA:
Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 LNSKIAFKIVSOEPA 15
Oy 1 LNSKIAFKIVSOEPA 15
RESULT 3
ID W78814 standard: peptide; 15 AA.
AC W78814:
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 190-204.
KW Microparticle: delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
OS Synthetic.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANCAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR WPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure: Page 8; 101pp: English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression

CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (1) a
CC fragment of a naturally-occurring mammalian protein; or (1i) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W7879;
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 15 AA:

Query Match 100.0%; Score 96; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 LNSKIAFKIVSOEPA 15
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 4
ID W07908 standard: protein; 614 AA.
AC W07908:
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHIKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto:antibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9pp: Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA:

Query Match 100.0%; Score 96; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 3.93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 189 LNSKIAFKIVSOEPA 203
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 5
ID R30742 standard: Protein; 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.

PN US798918-A.
PD 15-DEC-1992.
PE 27-NOV-1991: 798918.
PR 27-NOV-1991: US-798918.
PI (US\$H) US DEPT HEALTH & HUMAN SERVICE.
PA Amagel M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
DR N-PSDB: 035992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PS diagnostic and therapeutic uses.
PS Disclosure: Fig 7: 50pp: English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA:

Query Match 100.0%; Score 96; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 3,93e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 LNSKIAKIVSOEPA 204
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 6
ID W15489 standard; protein: 778 AA.
AC W15489;
DT 17-JUN-1997 (first entry)
DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.
FH Key 1.545
FT domain Location/Qualifiers
FT domain /note="Pemphigus foliaceus antigen protein"
FT domain
PD 25-MAR-1997.
PE 12-SEP-1995: 260899.
PR 12-SEP-1995: JP-260899.
PA (NISH/) NISHIKAWA T.
DR WPI: 97-241758/22.
DR P-PSDB: T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT through the hinge region used to treat pemphigus foliaceus
PS Claim 1: Page 10-12: 17pp: Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adsorbent upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is useful
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA:

Query Match 90.6%; Score 87; DB 1; Length 778;
Best Local Similarity 80.0%; Pred. No. 5,50e-03;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAKIVSOEPA 204
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 7
ID W13010 standard; protein: 263 AA.
AC W13010;

DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PE 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PI (PROG-) PROGEN BIOTECHNIK GMBH.
PA Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 9: Page 5; 8pp: German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinoma.
SQ Sequence 263 AA:

Query Match 80.2%; Score 77; DB 1; Length 263;
Best Local Similarity 73.3%; Pred. No. 9,68e-02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 41 LNSKISRYISLEPA 55
OY 1 LNSKIAKIVSOEPA 15
|||||

RESULT 8
ID W13009 standard; protein: 560 AA.
AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PE 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PI (PROG-) PROGEN BIOTECHNIK GMBH.
PA Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7: Page 5; 8pp: German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or

CC Carcinomas. 560 AA;
SO Sequence

Query Match 80.2%; Score 77; DB 1; Length 560;
Best Local Similarity 73.3%; Pred. No. 9,68e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 143 LNSKISRYVLSLEPA 157
1 LNSKIAFKIVSOEPA 15

RESULT 9
ID R55060 standard; Protein: 878 AA.
AC R55060:1994 (first entry)
DE Sequence of human liver E-cadherin.
KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
KW uvomorulin; L-CAM; Cell CAM 120/80.
OS Homo sapiens.
PN M09411401-A.
PD 26-MAY-1994.
PF 16-NOV-1993: U11097.
PR 17-NOV-1992: US-978897.
PA (UYVA) UNIV YALE.
PI Morrow JS, Rimm DL;
DR MPI: 94-183426/22.
N-PSDB: 065487.

PT Purified human E-cadherin protein and nucleic acid - used to develop prods. for diagnosis, prognosis, therapy and prophylaxis of E-cadherin disorders, e.g. malignancies
PS Claim 1: Page 59-63; 97pp; English.
CC E-cadherin is a cell adhesion molecule that is also known as uvomorulin. L-CAM and Cell CAM 120/80. The DNA encoding hEC was obtd. by screening normal human liver and hepatocellular carcinoma cDNA libraries and a colonic epithelial cell cDNA library. The following sequences are specifically claimed: AAs 1-878; 151-878; 30-401; AAs 402-513; AAs 178-878; AAs 1-150; AAs 178-289; AAs 290-401; AAs 402-513; AAs 151-703; AAs 1-703; AAs 728-878; AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-1648; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742; 1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from AAs 308-878. The prods. can be used in the diagnosis, prognosis, therapy and prophylaxis of conditions involving improper E-cadherin expression. Suitable dosages for i.v. admin. of a protein are 20-500 mcg/kg body wt.

SO Sequence 878 AA;

Query Match 56.3%; Score 54; DB 1; Length 878;
Best Local Similarity 53.8%; Pred. No. 4,96e+01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 293 NAAIVYTLISODP 305
1:11:1111:1
2 NSKIAFKIVSOEP 14

RESULT 10
ID R85487 standard; Protein: 878 AA.
AC R85487:
DT 18-MAR-1996 (first entry)
DE Human E-cadherin precursor.
KW E-cadherin; T-lymphocyte; alpha-E; beta-7 integrin; cell adhesion; auto-immune disease; Crohn disease; psoriasis.
OS Homo sapiens.
PN Key
FT peptide
FT 1. 150
FT location/Qualifiers
FT 151. 702
FT domain
FT /label- Sig-peptide
FT /label- Extracellular-domain
FT /note- "the extracellular domain (amino acids 1-552 of the mature protein) is the preferred

FT region for generation of peptides of the invention.
FT 703. 726
FT /label- Transmembrane-domain
FT domain 727. 876
FT /label- Cytoplasmic-domain

MO9529693-A1.
PD 09-NOV-1995.
PF 03-MAY-1995: U05518.
PR 03-MAY-1994: US-237919.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PI Brenner MB, Ceppek KL;
DR MPI: 95-392921/50.
N-PSDB: T05764.

PT Inhibiting adhesion of T lymphocytes with E-cadherin - useful for isolating agents to treat auto-immune diseases e.g. Crohn's disease, psoriasis, etc
PS Disclosure: Page 70-75; 103pp; English.
CC The human E-cadherin protein precursor (R85487) is expressed by an cDNA clone (T05764) derived from human liver. The extracellular domain of E-cadherin is used to generate peptides that specifically bind to heterotypic cognates of E-cadherin and which inhibit adhesion of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial or endothelial cells in vitro or in vivo, thereby modulating mucosal immune responses. Such peptides are also specifically reactive with a monoclonal antibody (E4.6 or E6.1) that binds to E-cadherin and that can inhibit T-cell binding.

SO Sequence 878 AA;

Query Match 56.3%; Score 54; DB 1; Length 878;
Best Local Similarity 53.8%; Pred. No. 4,96e+01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 293 NAAIVYTLISODP 305
1:11:1111:1
2 NSKIAFKIVSOEP 14

RESULT 11
ID W48711 standard; Protein: 2233 AA.
AC W48711:
DT 13-OCT-1998 (first entry)
DE HPrV-3 JS isolate wild-type L protein.
KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity; single stranded RNA virus; Mononegavirales.
OS Human parainfluenza virus.
PN W09813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997: U16718.
PR 27-SEP-1996: US-026823.
PA (AMCY) AMERICAN CYANAMID CO.
PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR MPI: 98-230710/20.
N-PSDB: V18272.

PT Recombinantly-generated, attenuated, non-segmented, negative-sense, single stranded RNA virus of order Mononegavirales - having attenuating mutation in 3' genomic promoter region and RNA polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure: Page 246-254; 420pp; English.
CC This sequence represents the wild-type L protein from Human parainfluenza virus (HPrV-3) type 3 isolate JS. This sequence is used in a method which involves the isolation of recombinantly-generated, attenuated, non-segmented, negative-sense, single stranded RNA virus of the order Mononegavirales which have at least 1 attenuating mutation in the 3' genomic promoter region and at least 1 attenuating mutation in the RNA polymerase gene. This RNA virus can be used as a vaccine to immunise an individual against such a virus.

SO Sequence 2233 AA;

Query Match 56.3%; Score 54; DB 1; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4,96e+01;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 12
 ID W48713 standard; Protein: 2233 AA.
 AC W48713;
 DT 13-OCT-1998 (first entry)
 DE HPIV-3 Vero cp45 vaccine L protein.
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KM single stranded RNA virus; Mononegavirales.
 OS Human parainfluenza virus.
 PN WO9813501-A2.
 PD 02-APR-1998.
 PF 19-SEP-1997; U16718.
 PR 27-SEP-1996; US-026823.
 PA (AMCY) AMERICAN CYANAMID CO.
 PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
 DR Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 WPI: 98-230710/20.
 N-PSDB: V18274.
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 PS Disclosure: Page 283-291; 426pp; English.
 CC This sequence represents the Human parainfluenza virus (HPIV-3) type 3
 CC vaccine Vero cp45 L protein. This sequence is used in a method which
 CC involves the isolation of recombinantly-generated, attenuated,
 CC non-segmented, negative-sense, single stranded RNA virus of the order
 CC Mononegavirales which have at least 1 attenuating mutation in the 3'
 CC genomic promoter region and at least 1 attenuating mutation in the RNA
 CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
 CC individual against such a virus.
 SQ Sequence 2233 AA;

Query Match 56.3%; Score 54; DB 1; Length 2233;
 Best Local Similarity 33.3%; Pred. No. 4.96e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 13
 ID W48712 standard; Protein: 2233 AA.
 AC W48712;
 DT 13-OCT-1998 (first entry)
 DE HPIV-3 FRh1 cp45 vaccine L protein.
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KM single stranded RNA virus; Mononegavirales.
 OS Human parainfluenza virus.
 PN WO9813501-A2.
 PD 02-APR-1998.
 PF 19-SEP-1997; U16718.
 PR 27-SEP-1996; US-026823.
 PA (AMCY) AMERICAN CYANAMID CO.
 PI (USSH) US DEPT HEALTH & HUMAN SERVICES.
 DR Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 WPI: 98-230710/20.
 N-PSDB: V18273.
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 PS Disclosure: Page 265-273; 426pp; English.
 CC This sequence represents the Human parainfluenza virus (HPIV-3) type 3
 CC vaccine FRh1 cp45 L protein. This sequence is used in a method which
 CC involves the isolation of recombinantly-generated, attenuated,
 CC non-segmented, negative-sense, single stranded RNA virus of the order

CC Mononegavirales which have at least 1 attenuating mutation in the 3'
 CC genomic promoter region and at least 1 attenuating mutation in the RNA
 CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
 CC individual against such a virus.
 SQ Sequence 2233 AA;

Query Match 56.3%; Score 54; DB 1; Length 2233;
 Best Local Similarity 33.3%; Pred. No. 4.96e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDRSVLYRINNOEPG 987
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 14
 ID W82841 standard; Protein: 456 AA.
 AC W82841;
 DT 03-FEB-1999 (first entry)
 DE Human cerebral protein-1
 KW Human; cerebral protein-1; HUCEP-1; Ischaemic brain disease;
 KW neuron function activating activity; nerve denatured disease;
 KW Parkinson's disease; Alzheimer's disease.
 OS Homo sapiens.
 PN J10257891-A.
 PD 29-SEP-1998.
 PF 19-MAR-1997; 065716.
 PR 19-MAR-1997; JP-065716.
 PA (TAIS) TAISHO PHARM CO LTD.
 DR WPI: 98-575902/49.
 N-PSDB: V64162.
 PT HUCEP-1 protein having neuron function activating activity - useful
 PT for treating ischemic brain diseases and nerve denaturation
 PT conditions such as Parkinson's and Alzheimer's diseases
 PS Claim 1; Fig 4; 17pp; Japanese.
 CC The present sequence represents human cerebral protein-1 (HUCEP-1).
 CC HUCEP-1 has neuron function activating activity. HUCEP-1 is useful
 CC for treating ischemic brain diseases and nerve denaturation
 CC conditions such as Parkinson's and Alzheimer's diseases.
 SQ Sequence 456 AA;

Query Match 55.2%; Score 53; DB 1; Length 456;
 Best Local Similarity 54.5%; Pred. No. 6.41e+01;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 408 TKUGFKIVSKD 418
 Oy 3 SKIAFKIVSOE 13

RESULT 15
 ID R40843 standard; Protein: 572 AA.
 AC R40843;
 DT 24-FEB-1994 (first entry)
 DE Bilirubin oxidase.
 KW Bilirubin; oxidase; analytical; BO; expression vector; PCR;
 KM polymerase chain reaction.
 OS Myrothecium verrucaria.
 FH key
 FT Location/Qualifiers
 FT peptide 1..38
 FT /label= sig_peptide
 FT protein 39..534
 FT /label= mat_protein
 PN J05199882-A.
 PD 10-AUG-1993.
 PF 24-JAN-1992; 034126.
 PR 24-JAN-1992; JP-034126.
 PA (AMAN) AMANO PHARM KK.
 DR WPI: 93-284681/36.
 N-PSDB: 047790.
 PT Bilirubin oxidase prep. useful as an analytical enzyme - by
 PT culturing bilirubin oxidase in transformant culture
 PS Claim 1; Page 29-32; 32pp; Japanese.

Mon May 15 08:08:21 2000

US-08-991-628-3a.rag

Page 6

CC The sequence encodes bilirubin oxidase. The protein produced has a
CC 38 amino acid signal peptide which is removed to give the mature
CC protein (R40843).
CC Sequence 572 AA;
SQ

```
Query Match      51.0%; Score 49; DB 1; Length 572;
Best Local Similarity 13.3%; Pred. NO. 1.75e+02;
Matches      2; Conservative 12; Mismatches 1; Indels 0; Gaps 0;
```

```
Db      282 IDTRLPEKVIASDSG 296
          :::::|::: :::
OY      1 LNSKIAFKIVSQEPA 15
```

Search completed: Sat May 13 07:10:55 2000
Job time : 8 secs.

MUSE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:10:21 2000; Maspar time 4.28 Seconds
Tabular output not generated. 165.192 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.750; Variance 32.305; scale 0.797

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	96	100.0	999	1	IJHUG3	desmoglein 3 precursor 1.15e-09
2	87	90.6	1043	1	IJBOG1	desmoglein 1 precursor 2.27e-07
3	87	90.6	1049	1	IJHUG1	desmoglein 1 precursor 2.27e-07
4	77	80.2	1117	2	S38673	desmoglein 2 - human 6.37e-05
5	61	63.5	1421	2	T02501	hypothetical protein 2.63e-01
6	60	62.5	884	1	IJM5CE	E-cadherin precursor 4.27e-01
7	60	62.5	884	1	IJM5CE	uvomorulin - mouse 4.27e-01
8	60	62.5	906	1	IJHUCN	cadherin 2 precursor 6.90e-01
9	59	61.5	385	2	S49752	homeotic protein YOX1 1.11e+00
10	58	60.4	138	2	S74520	probable membrane protein 1.77e+00
11	57	59.4	1264	2	S64146	cadherin-related tumor 4.45e+00
12	57	59.4	5147	1	IJFFTM	translation elongation 7.00e+00
13	55	57.3	192	2	F70126	N-cadherin precursor 7.00e+00
14	54	56.3	108	2	T05921	N-cadherin precursor 7.00e+00
15	54	56.3	877	1	IJBOCN	cadherin 1 precursor 7.00e+00
16	54	56.3	882	1	IJHUCE	cadherin 1 precursor 7.00e+00
17	54	56.3	905	2	S43064	N-cadherin precursor 7.00e+00
18	54	56.3	906	1	IJM5CN	cadherin 1 precursor 7.00e+00
19	54	56.3	1146	2	S64402	probable transmembrane protein 1.77e+00
20	54	56.3	2233	1	IJHUCN	cadherin 2 precursor 6.90e-01
21	53	55.2	775	2	T07172	subtilisin-like protease 1.09e+01
22	53	55.2	887	1	IJHUCN	cadherin 1 precursor 7.00e+00
23	53	55.2	2048	1	ZLNZSE	genome polypeptide 1.09e+01

24	53	55.2	2228	1	ZLNZSV	genome polypeptide 1.09e+01
25	52	54.2	191	2	C72411	holiday junction DNA 1.70e+01
26	52	54.2	485	2	A17266	probable transcriptio 1.70e+01
27	52	54.2	732	1	IJCHCB	B-cadherin precursor 1.70e+01
28	52	54.2	821	1	B34488	calpain (EC 3.4.22.17 1.70e+01
29	51	53.1	241	2	S63634	ribosomal protein S3 2.62e+01
30	51	53.1	293	2	F49846	spa33 protein - Shige 2.62e+01
31	51	53.1	293	2	E42284	spa33 protein - Shige 2.62e+01
32	51	53.1	644	2	G64938	hypothetical protein 2.62e+01
33	51	53.1	742	2	A49341	isocitrate dehydrogen 2.62e+01
34	51	53.1	813	3	T02672	hypothetical protein 2.62e+01
35	51	53.1	947	2	D72067	polymorphic membrane 2.62e+01
36	51	53.1	2764	2	T13949	neurofibromin - fruit 2.62e+01
37	51	53.1	2802	2	T13945	neurofibromin - fruit 2.62e+01
38	51	53.1	2802	2	T13947	neurofibromin - fruit 2.62e+01
39	50	52.1	217	2	S25314	aspartic proteinase 1 4.01e+01
40	50	52.1	277	2	JC5284	carboxyl reductase (N 4.01e+01
41	50	52.1	355	2	T15203	hypothetical protein 4.01e+01
42	50	52.1	461	2	A71662	sodium/pantothenate 4.01e+01
43	50	52.1	458	2	A71662	hypothetical protein 4.01e+01
44	50	52.1	578	2	T02292	hypothetical protein 4.01e+01
45	50	52.1	1045	2	T13481	genome polypeptide 1 4.01e+01
			2410	1	J01948	genome polypeptide 1 4.01e+01

ALIGNMENTS

RESULT 1
ENTRY IJHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal name Homo sapiens #common name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 23-Jun-1999

ACCESSIONS
REFERENCE A41088
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references GDB:M76483; NID:9190751; PID:AA60230.1; PID:9190751
#accession A41088
#molecule_type mRNA
#residues 1-999 #label AMA
#cross-references GDB:M76483; NID:9190751; PID:AA60230.1; PID:9190751

GENETICS

#gene GDB:DSG3

#cross-references GDB:134030; OMIM:169615

#map_position 18q12.1-18q12.2

CLASSIFICATION #superfamily cadherin; cadherin repeat homology

KEYWORDS #binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE #domain signal sequence #status predicted #label SIG

SUMMARY #length 999 #molecular-weight 107502 #checksum 8311

Query Match 100.0%; Score 96; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.15e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ENTRY 1JB0G1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES #formal_name Bos primigenius taurus #common_name cattle
 ORGANISM 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
 DATE 22-Jun-1999

ACCESSIONS
 REFERENCE S14603: A38872: A37785: S38721: A48173: S24412
 S14603
 #authors Koch, P.J.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.
 #submission Submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 #label KOC
 REFERENCE #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 A38872
 #authors Koch, P.J.: Goldschmidt, M.D.: Walsh, M.J.: Zimbelmann, R.: Franke, W.W.

#journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references WUID:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87:968-1043 #label KO2
 REFERENCE #cross-references GB:S64268; GB:S64270
 A37785

#authors Goodwin, L.: Hill, J.E.: Raynor, K.: Raszi, L.: Manabe, M.: Cowlin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references WUID:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123,'V',125-493 #label GOO
 REFERENCE #cross-references GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
 S38721

#authors Zimbelmann, R.
 #submission Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 #label ZIM
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062

REFERENCE A48173
 #authors Koch, P.J.: Walsh, M.J.: Schmelz, M.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references WUID:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001,'AQP5AT' #label KO3
 REFERENCE #cross-references GB:X57784
 S38721

#note this sequence has been revised in references A38872 and S38721

GENETICS
 #gene DSG1
 #classification #superfamily cadherin: cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein;

FEATURE
 1-22
 24-49
 50-1043
 50-548
 52-157
 150-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180,496

transmembrane protein

SUMMARY
 #length 1043 #molecular_weight 112242 #checksum 6897

Query Match
 Best Local Similarity 80.0% Pred. No. 2,27e-07;
 Matches 12: Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ENTRY 1IHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
 22-Jun-1999

ACCESSIONS
 REFERENCE S16906: A39706: A61254: S16158
 S16906
 #authors Buxton, R.S.
 #submission Submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 #label BUX
 REFERENCE #cross-references EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
 A39706

#authors Wheeler, G.N.: Parker, A.E.: Thomas, C.L.: Attalio, P.: Poynter, D.: Arnemann, J.: Rutman, A.J.: Pidsley, S.C.: Watt, F.M.: Rees, D.A.: Buxton, R.S.: Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4756-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references WUID:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 #label WHE
 REFERENCE #cross-references GB:X56654
 A61254

#authors Nilles, L.A.: Parry, D.A.D.: Powers, E.E.: Angst, B.D.: Wagner, R.M.: Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references WUID:92121251
 #accession A61254
 #molecule_type mRNA
 #residues 26-1049 #label NIL

REFERENCE A61279
 #authors Wheeler, G.N.: Buxton, R.S.: Parker, A.E.: Arnemann, J.: Rees, D.A.: King, I.A.: Magee, A.I.

#journal Biochem. Soc. Trans. (1991) 19:1060-1064
#title Desmosomal glycoproteins I, II and III: novel members of the
#crosstalk superfamily.
#crosstalk-references MUID:92175187
#accession A61279
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-55 ##label WH3

GENETICS
#gene GDB:DSG1
#crosstalk-references GDB:126563; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin: cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG
24-49 #domain propeptide #status predicted #label PRO
50-1049 #product desmoglein #status predicted #label MAR
50-548 #domain extracellular #status predicted #label EXT
52-157 #domain cadherin repeat homology #label CR1
160-269 #domain cadherin repeat homology #label CR2
272-385 #domain cadherin repeat homology #label CR3
392-493 #domain cadherin repeat homology #label CR4
509-530 #region serine/threonine-rich
549-569 #domain transmembrane #status predicted #label TMN
572-1049 #domain intracellular #status predicted #label INT
840-869 #domain desmoglein repeat #label DG1
870-899 #domain desmoglein repeat #label DG2
900-927 #domain desmoglein repeat #label DG3
928-956 #domain desmoglein repeat #label DG4
969-1019 #region glycine/serine-rich
110,180 #binding_site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY
#length 1049 #molecular_weight 113715 #checksum 4482

Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 2,27e-07;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIIVSOEPA 204
OY 1 LNSKIAFKIIVSOEPA 15

RESULT 4
ENTRY S38673 #type complete
TITLE desmoglein 2 - human
ALTERNATE_NAMES #formal_name Homo sapiens #common_name man
ORGANISM 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
DATE 20-Aug-1999
ACCESSIONS S38673; B38672
REFERENCE S38673
#authors Zimbelmann, R.
#submission submitted to the EMBL Data Library, September 1993
#accession S38673
#status preliminary
#molecule_type mRNA
#residues 1-1117 ##label ZIM
#crosstalk-references EMBL:Z26317; NID:9416177; PIDN:CAA81226.1;
PID:9416178

REFERENCE A38872
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.;
Frank, W.W.
#journal Eur. J. Cell Biol. (1991) 55:200-208
#title Complete amino acid sequence of the epidermal desmoglein
precursor polypeptide and identification of a second type
of desmoglein gene.
#crosstalk-references MUID:92037656
#accession B38872
#molecule_type mRNA
#residues 777-1117 ##label KOC

##crosstalk-references GB:S64273

GENETICS
#gene GDB:DSG2
#crosstalk-references GDB:128808; OMIM:125671
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin: cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
membrane protein

FEATURE
51-158 #domain cadherin repeat homology #label CR1
161-271 #domain cadherin repeat homology #label CR2
SUMMARY #length 1117 #molecular_weight 122384 #checksum 7660

Query Match 80.2%; Score 77; DB 2; Length 1117;
Best Local Similarity 73.3%; Pred. No. 6,37e-05;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 191 LNSKISYRIVSLEPA 205
OY 1 LNSKIAFKIIVSOEPA 15

RESULT 5
ENTRY T02501 #type complete
TITLE hypothetical protein T19C21.7 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02501
REFERENCE 214676
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.W.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC T19C21 genomic
sequence.
#accession T02501
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-1421 ##label ROU
#crosstalk-references EMBL:AC004683; NID:93395421; PID:93395428
##experimental_source cultivar Columbia

GENETICS
#map_position 2
#introns 52/2; 107/1; 148/2; 191/3; 212/2; 1317/3; 1346/3
#note T19C21.7
SUMMARY #length 1421 #molecular_weight 154325 #checksum 9888

Query Match 63.5%; Score 61; DB 2; Length 1421;
Best Local Similarity 40.0%; Pred. No. 2,63e-01;
Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Db 262 VDSKISYEITONPG 276
OY 1 LNSKIAFKIIVSOEPA 15

RESULT 6
ENTRY IUMSC2 #type complete
TITLE E-cadherin precursor, epithelial - mouse
ALTERNATE_NAMES uvomorulin
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
ACCESSIONS S04528; S03160; I49565; S48735
REFERENCE S04528
#authors Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.;
Takeichi, M.
#journal Nature (1987) 329:341-343
#title Transformation of cell adhesion properties by exogenously
introduced E-cadherin cDNA.

#cross-references MWID:87315445
#accession S04528
#molecule-type mRNA
#residues 1-412,'V',414-884 #label NAG
#cross-references EMBL:X06115
REFERENCE S03160
#authors Ringwald, M.; Schuh, R.; Vestweber, D.; Elstetter, H.;
Lottspeich, F.; Engel, J.; Doebe, R.; Jaehnig, F.; Eppien,
J.; Mayer, S.; Mueller, C.; Kemler, R.
#journal EMBO J. (1987) 6:3647-3653
#title The structure of cell adhesion molecule uvomorulin. Insights
into the molecular mechanism of Ca(2+)-dependent cell
adhesion.
#cross-references MWID:88111553
#accession S03160
#molecule-type mRNA
#residues 157-884 #label RIN
#cross-references EMBL:X06339
#note mature protein, was confirmed by protein sequencing
part of this sequence, including the amino end of the
REFERENCE I49565
#authors Behrens, J.; Loewick, O.; Klein-Hitpass, L.; Birchmeier, W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:11495-11499
#title The E-cadherin promoter: Functional analysis of a G-C-rich
region and an epithelial cell-specific palindromic
regulatory element.
#cross-references MWID:92107977
#accession I49565
#status preliminary: translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 1-15 #label RES
#cross-references GB:M01449; NID:g192325; PIDN:AAA37352.1; PID:g192326
REFERENCE S48735
#authors Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Potumb, T.;
Takeichi, M.; Ikura, M.
#journal FEBS Lett. (1994) 352:318-322
#title Purification and spectroscopic characterization of a
recombinant amino-terminal polypeptide fragment of mouse
epithelial cadherin.
#cross-references MWID:95010732
#accession S48735
#status preliminary
#molecule-type protein
#residues 156-300 #label TON
COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are
thought to be involved in the sorting of different cell types
during morphogenesis.
GENETICS
#gene E-cadherin
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein
FEATURE
1-77 #domain signal sequence #status predicted #label SIG\
28156 #domain propeptide #status predicted #label PRO\
157-884 #product E-cadherin, epithelial #status experimental
#label MAT\
157-699 #domain extracellular #status predicted #label EXT\
159-264 #domain cadherin repeat homology #label CR1\
234-239 #region cadherin binding #status predicted\
267-377 #domain cadherin repeat homology #label CR2\
380-488 #domain cadherin repeat homology #label CR3\
489-597 #domain cadherin repeat homology #label CR4\
598-702 #domain cadherin repeat homology #label CR5\
703-733 #domain transmembrane #status predicted #label TMN\
734-884 #domain intracellular #status predicted #label INT\
842-885 #region serine-rich\
560,659 #binding-site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 884 #molecular-weight 98255 #checksum 2268
Query Match 62.5%; Score 60; DB 1; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.27e-01;

Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:
DB 299 NAAIATVSDP 311
1: 11: 1111:1
Qy 2 NSKIARIVSOEP 14
RESULT 7
ENTRY S34438 #type complete
TITLE uvomorulin - mouse
ORANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change
23-May-1997
ACCESSIONS S34438
REFERENCE S34438
#authors Ringwald, M.; Barbaulet, H.; Schmidt, C.; Kemler, R.
#journal Nucleic Acids Res. (1991) 19:6533-6539
#title The structure of the gene coding for the mouse cell adhesion
molecule uvomorulin.
#cross-references MWID:92093614
#accession S34438
#status preliminary: nucleic acid sequence not shown:
translation not shown
#molecule-type DNA
#residues 1-884 #label RIN
#cross-references EMBL:X60975
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1992
GENETICS
#introns 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3;
524/2; 573/1; 648/1; 724/1; 767/3; 815/3
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
FEATURE
380-488 #domain cadherin repeat homology #label CR3
SUMMARY #length 884 #molecular-weight 98283 #checksum 2125
Query Match 62.5%; Score 60; DB 2; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.27e-01;
Matches 8: Conservative 3: Mismatches 2: Indels 0: Gaps 0:
DB 299 NAAIATVSDP 311
1: 11: 1111:1
Qy 2 NSKIARIVSOEP 14
RESULT 8
ENTRY ITHUCN #type complete
TITLE cadherin 2 precursor - human
ALTERNATE_NAMES N-cadherin; neuronal cadherin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
22-Jun-1999
ACCESSIONS A38870; S11487; J00751; S13799
REFERENCE A38870
#authors Reid, R.A.
#submissions submitted to the EMBL Data Library, November 1990
#accession A38870
#molecule-type mRNA
#residues 1-906 #label REI
#cross-references EMBL:X54315; NID:g34998; PIDN:CMA38213.1; PID:g34999
REFERENCE S11487
#authors Reid, R.A.; Hemperly, J.J.
#journal Nucleic Acids Res. (1990) 18:5896
#title Human N-cadherin: nucleotide and deduced amino acid sequence.
#cross-references MWID:91016546
#accession S11487
#molecule-type mRNA
#residues 1-340,'N',342-698,'R',700-704,'F',706-906 #label RE2
#cross-references EMBL:X54315
#note this sequence has been revised in reference A38870
REFERENCE J00751
#authors Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell,
D.; Spurr, N.; Goodfellow, P.N.

##cross-references EMBL:X62392; NID:g5501; PIDN:CAA44264.1; PID:g5502

GENETICS
#gene SCD:YOX1
##cross-references SCD:S0004489; MIPS:YML027w

CLASSIFICATION
#map_position 13L
#superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation

FEATURE
177-233
SUMMARY #domain homeobox homology #label HOX
#length 385 #molecular-weight 42739 #checksum 6962

Query Match 61.5% Score 59; DB 2; Length 385;
Best Local Similarity 63.6% Pred. No. 6.90e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 364 KFGKIVDQDP 374
Oy 4 KIAFKIVSQEP 14

##cross-references EMBL:246559; NID:g575680; PIDN:CAA66628.1;
PID:g575692; MIPS:YML027w

GENETICS
#gene GDB:CDH2; NCAD
##cross-references GDB:M34064

CLASSIFICATION
#map_position 18q12.1-18q12.1
#superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-27
28-159
160-906
160-714
162-267
237-242
270-382
385-497
500-605
606-712
715-746
747-906
865-878
190,273,325,402,
572,622,651,692

#domain signal sequence #status predicted #label SIG
#domain propeptide #status predicted #label PRO
#product N-cadherin #status predicted #label MAT
#domain extracellular #status predicted #label EXT
#domain cadherin repeat homology #label CR1
#domain cadherin repeat homology #label CR2
#domain cadherin repeat homology #label CR3
#domain cadherin repeat homology #label CR4
#domain cadherin repeat homology #label CR5
#domain transmembrane #status predicted #label TM1
#domain intracellular #status predicted #label INT
#region serine-rich
#binding-site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 906 #molecular-weight 99864 #checksum 9191

Query Match 62.5% Score 60; DB 1; Length 906;
Best Local Similarity 46.7% Pred. No. 4.27e-01;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 300 LNCALRYRIVSQAPS 314
Oy 1 LNSKIAFKIVSQEPA 15

RESULT 9
ENTRY S49752 #type complete
TITLE homeotic protein YOX1 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES #formal_name YML027w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 24-Sep-1999

ACCESSIONS S49752; S33388
REFERENCE S49741
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, November 1994
#accession S49752
##molecule-type DNA
##residues 1-385 #label BAD
##cross-references EMBL:246559; NID:g575680; PIDN:CAA66628.1;
PID:g575692; MIPS:YML027w

REFERENCE S33388
#authors Kaufmann, E.
#journal Chromosome (1993) 102:174-179
#title In vitro binding to the leucine tRNA gene identifies a novel yeast homeobox gene.
#cross-references WUID:93209080
#accession S33388
##molecule-type DNA
##residues 1-290, 'OGLITP' #label KAU

##cross-references EMBL:X62392; NID:g5501; PIDN:CAA44264.1; PID:g5502

GENETICS
#gene SCD:YOX1
##cross-references SCD:S0004489; MIPS:YML027w

CLASSIFICATION
#map_position 13L
#superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation

FEATURE
177-233
SUMMARY #domain homeobox homology #label HOX
#length 385 #molecular-weight 42739 #checksum 6962

Query Match 61.5% Score 59; DB 2; Length 385;
Best Local Similarity 63.6% Pred. No. 6.90e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 364 KFGKIVDQDP 374
Oy 4 KIAFKIVSQEP 14

##cross-references EMBL:246559; NID:g575680; PIDN:CAA66628.1;
PID:g575692; MIPS:YML027w

GENETICS
#gene GDB:CDH2; NCAD
##cross-references GDB:M34064

CLASSIFICATION
#map_position 18q12.1-18q12.1
#superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-27
28-159
160-906
160-714
162-267
237-242
270-382
385-497
500-605
606-712
715-746
747-906
865-878
190,273,325,402,
572,622,651,692

#domain signal sequence #status predicted #label SIG
#domain propeptide #status predicted #label PRO
#product N-cadherin #status predicted #label MAT
#domain extracellular #status predicted #label EXT
#domain cadherin repeat homology #label CR1
#domain cadherin repeat homology #label CR2
#domain cadherin repeat homology #label CR3
#domain cadherin repeat homology #label CR4
#domain cadherin repeat homology #label CR5
#domain transmembrane #status predicted #label TM1
#domain intracellular #status predicted #label INT
#region serine-rich
#binding-site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 906 #molecular-weight 99864 #checksum 9191

Query Match 62.5% Score 60; DB 1; Length 906;
Best Local Similarity 46.7% Pred. No. 4.27e-01;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 300 LNCALRYRIVSQAPS 314
Oy 1 LNSKIAFKIVSQEPA 15

RESULT 9
ENTRY S49752 #type complete
TITLE homeotic protein YOX1 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES #formal_name YML027w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 24-Sep-1999

ACCESSIONS S49752; S33388
REFERENCE S49741
#authors Badcock, K.; Churcher, C.
#submission submitted to the EMBL Data Library, November 1994
#accession S49752
##molecule-type DNA
##residues 1-385 #label BAD
##cross-references EMBL:246559; NID:g575680; PIDN:CAA66628.1;
PID:g575692; MIPS:YML027w

REFERENCE S33388
#authors Kaufmann, E.
#journal Chromosome (1993) 102:174-179
#title In vitro binding to the leucine tRNA gene identifies a novel yeast homeobox gene.
#cross-references WUID:93209080
#accession S33388
##molecule-type DNA
##residues 1-290, 'OGLITP' #label KAU

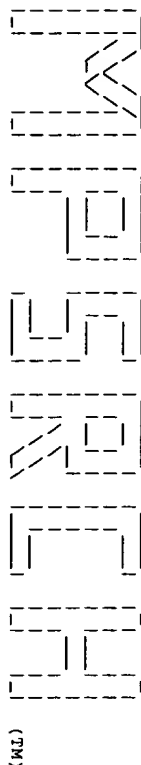
REFERENCE 564144
#authors Escribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
#submission Submitted to the Protein Sequence Database, May 1996
#accession 564146
#molecule_type DNA
#residues 1-1264 #label ESC
#cross-references EMBL:272655; NID:g1322701; PID:e243361; PID:g1322702;
MIPS:YGL133W
#experimental_source strain S288C
REFERENCE
#authors Escribano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
#journal Yeast (1996) 12:887-892
#title Sequence analysis of a 14.6 kb DNA fragment of *Saccharomyces cerevisiae* chromosome VII reveals SEC27, SSM1b, a putative S-adenosylmethionine-dependent enzyme and six new open reading frames.
#cross-references MUID:96437978
#accession S71739
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-1264 #label ESM
#cross-references EMBL:X82670
#note the nucleotide sequence was submitted to the EMBL Data Library, October 1995
GENETICS
#map_position 7L
KEYWORDS transmembrane protein
FEATURE
349-565 #length 1264 #molecular-weight 145642 #checksum 9929
SUMMARY
Query Match 59.4%; Score 57; DB 2; Length 1264;
Best Local Similarity 42.9%; Pred. No. 1.77e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
DB 524 SDRVEKYIHDDPA 537
QY 2 NSKIAFKIVSOEPA 15
RESULT 12
ENTRY IUFFTM #type complete
TITLE cadherin-related tumor suppressor precursor - fruit fly
ORGANISM *Drosophila melanogaster*
#formal_name *Drosophila melanogaster*
30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
ACCESSION A41087
REFERENCE
#authors Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Blessmann, H.; Bryant, P.J.; Goodman, C.S.
#journal Cell (1991) 67:853-868
#title The fat tumor suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily.
#cross-references MUID:92069752
#accession A41087
#molecule_type mRNA
#residues 143-485;1279-5147 #label MAH
#cross-references GB:M80537
#accession B41087
#molecule_type DNA
#residues 1-142;487-1278 #label MAZ
#cross-references GB:M80537
#note 1229-Gly and 1233-Ser were also found
GENETICS
#gene fat
#cross-references FlyBase:Fbgn0001075
CLASSIFICATION #superfamily cadherin-related tumor suppressor; cadherin
KEYWORDS repeat homology; EGF homology
calcium binding; cell adhesion; duplication; transmembrane protein
FEATURE
1-35 #domain signal sequence #status predicted #label SIG\

36-5147
36-4583
51-156
159-270
271-382
390-494
467-569
602-708
718-822
831-942
948-1049
1052-1153
1156-1278
1281-1384
1387-1489
1492-1601
1607-1713
1717-1823
1826-1922
1925-2027
2028-2167
2169-2278
2281-2384
2387-2491
2494-2596
2599-2703
2707-2810
2813-2913
2915-3013
3014-3124
3127-3229
3232-3334
3337-3439
3442-3545
3548-3651
3654-3756
3954-4010
4017-4048
4056-4089
4096-4127
4384-4609
4610-5147
#product cadherin-related tumor suppressor #status predicted #label MAY
#domain extracellular #status predicted #label EXT
#domain cadherin repeat homology #label CR1
#domain cadherin repeat homology #label CR2
#domain cadherin repeat homology #label CR3
#domain cadherin repeat homology #label CR4
#domain cadherin repeat homology #label CR5
#domain cadherin repeat homology #label CR6
#domain cadherin repeat homology #label CR7
#domain cadherin repeat homology #label CR8
#domain cadherin repeat homology #label CR9
#domain cadherin repeat homology #label CR10
#domain cadherin repeat homology #label CR11
#domain cadherin repeat homology #label CR12
#domain cadherin repeat homology #label CR13
#domain cadherin repeat homology #label CR14
#domain cadherin repeat homology #label CR15
#domain cadherin repeat homology #label CR16
#domain cadherin repeat homology #label CR17
#domain cadherin repeat homology #label CR18
#domain cadherin repeat homology #label CR19
#domain cadherin repeat homology #label CR20
#domain cadherin repeat homology #label CR21
#domain cadherin repeat homology #label CR22
#domain cadherin repeat homology #label CR23
#domain cadherin repeat homology #label CR24
#domain cadherin repeat homology #label CR25
#domain cadherin repeat homology #label CR26
#domain cadherin repeat homology #label CR27
#domain cadherin repeat homology #label CR28
#domain cadherin repeat homology #label CR29
#domain cadherin repeat homology #label CR30
#domain cadherin repeat homology #label CR31
#domain cadherin repeat homology #label CR32
#domain cadherin repeat homology #label CR33
#domain EGF homology #label EG1
#domain EGF homology #label EG2
#domain EGF homology #label EG3
#domain EGF homology #label EG4
#domain transmembrane #status predicted #label TM
#domain intracellular #status predicted #label INT
#length 5147 #molecular-weight 564895 #checksum 6994
SUMMARY
Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 1.77e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
DB 1520 LNKVSYAISKEP 1533
QY 1 LNSKIAFKIVSOEP 14
RESULT 13
ENTRY F70126 #type complete
TITLE translation elongation factor P (efp) homolog - Lyme disease spirochete
ORGANISM *Borrelia burgdorferi* #common_name Lyme disease spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
ACCESSION F70126
REFERENCE F70126
#authors Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Watthey, L.; McDonald, L.; Artach, P.; Boman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.

#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references MUID:98065943
#accession F70126
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-192 #label KLE
##cross-references GB:AE001132; GB:AE000783; NID:g2688107; PIDN:AAC66610.1; PID:g2688116; TIGR:BB0214
##experimental_source strain B31
CLASSIFICATION #superfamily translation elongation factor EF-P
SUMMARY #length 192 #molecular_weight 21413 #checksum 4881
Query Match 57.3%; Score 55; DB 2; Length 192;
Best Local Similarity 53.3%; Pred. No. 4,45e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 130 LAPKAFVEVEEA 144
OY 1 LNSKIAFKIVSOEPA 15
RESULT 14
ENTRY T05921 #type fragment
TITLE N-carbamyl-L-amino acid amidohydrolase homolog - barley (fragment)
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
T05921
215411
#authors Hess, W.R.; Golz, R.R.; Boerner, T.
#journal Plant Sci. (1998) 133:191-201
#title Analysis of randomly selected cDNAs reveals the expression of stress- and defence related genes in the barley mutant alosterians.
#accession T05921
#status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-108 #label HES
##cross-references EMBL:A222776; NID:e1203983; PID:e1203984
##experimental_source cv. Halsa, leaf
SUMMARY #length 108 #checksum 4694
Query Match 56.3%; Score 54; DB 2; Length 108;
Best Local Similarity 45.5%; Pred. No. 7,00e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 15 LEFKINODPP 25
OY 5 IAFKIVSOEPA 15
RESULT 15
ENTRY IJBOCN #type fragment
TITLE N-cadherin precursor - bovine (fragment)
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
I11693
S11693
S11693
#authors Liaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.
#journal EMBO J. (1990) 9:2701-2708
#title Identification and cloning of two species of cadherins in bovine endothelial cells.
#cross-references MUID:90360979
#accession S11693
##molecule_type mRNA
##residues 1-877 #label LIA
##cross-references EMBL:X53615; NID:g164; PIDN:CAA37677.1; PID:g664894

COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in the sorting of different cell types during morphogenesis.
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
FEATURE
1-130
#domain propeptide (fragment) #status predicted #label PRO
131-877
131-885 #product N-cadherin #status predicted #label MAT
133-238 #domain extracellular #status predicted #label EXT
208-213 #domain cadherin repeat homology #label CR1
241-253 #region cadherin binding #status predicted
356-468 #domain cadherin repeat homology #label CR2
471-576 #domain cadherin repeat homology #label CR3
577-685 #domain cadherin repeat homology #label CR4
686-717 #domain cadherin repeat homology #label CR5
718-877 #domain transmembrane #status predicted #label TM1
836-849 #domain intracellular #status predicted #label INT
161,244,296,373, #binding_site carbohydrate (Asn) (covalent) #status
543,593,622,663 predicted
SUMMARY #length 877 #checksum 8685
Query Match 56.3%; Score 54; DB 1; Length 877;
Best Local Similarity 40.0%; Pred. No. 7,00e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Db 271 LNCMLRYRIISQAPS 285
OY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 07:10:30 2000
Job time : 9 secs.


 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSEARCH protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:05:41 2000; Maspar time: 79.42 seconds

Tabular output not generated. 5.752 Million cell updates/sec

Title: >US-08-991-628-3

Description: (1-15) from US08991628.pep

Perfect Score: 96

Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150

Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

swiss-prot38
 1:swissprot

Statistics: Mean 26.327; Variance 28.416; scale 0.926

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	96	100.0	999	1	DSG3_HUMAN	DESMOGLEIN 3 PRECURSOR	1.17e-11
2	87	90.6	1043	1	DSG1_BOVIN	DESMOGLEIN 1 PRECURSOR	5.08e-09
3	87	90.6	1049	1	DSG1_HUMAN	DESMOGLEIN 1 PRECURSOR	5.08e-09
4	77	80.2	1117	1	DSG3_HUMAN	DESMOGLEIN 2 PRECURSOR	3.24e-06
5	60	62.5	884	1	CAD1_MOUSE	EPITHELIAL-CADHERIN PR	7.53e-02
6	60	62.5	906	1	CAD2_HUMAN	NEURAL-CADHERIN PRECUR	7.53e-02
7	59	61.5	385	1	YOX1_YEAST	HOMEOBOX PROTEIN YOX1	1.30e-01
8	57	59.4	1264	1	YGN3_YEAST	HYPOTHETICAL 145.6 KD	3.79e-01
9	57	59.4	5147	1	FAT_DROME	CADHERIN-RELATED TUMOR	3.79e-01
10	55	57.3	192	1	EPF_BOBBU	ELONGATION FACTOR P (E	1.08e+00
11	54	56.3	877	1	CAD2_BOVIN	NEURAL-CADHERIN PRECUR	1.79e+00
12	54	56.3	882	1	CAD1_HUMAN	EPITHELIAL-CADHERIN PR	1.79e+00
13	54	56.3	905	1	CAD6_XENLA	BLASTOMERE-CADHERIN PR	1.79e+00
14	54	56.3	906	1	CAD2_MOUSE	NEURAL-CADHERIN PRECUR	1.79e+00
15	54	56.3	1146	1	AS10_YEAST	ASK10 PROTEIN	1.79e+00
16	54	56.3	2233	1	RRPL_P13H4	RNA POLYMERASE BETA SU	2.97e+00
17	53	55.2	887	1	CAD1_CHICK	EPITHELIAL-CADHERIN PR	2.97e+00
18	53	55.2	2048	1	RRPL_SENDE	RNA POLYMERASE BETA SU	2.97e+00
19	53	55.2	2228	1	RRPL_SENDE	RNA POLYMERASE BETA SU	2.97e+00
20	53	55.2	2228	1	RRPL_SENDE	RNA POLYMERASE BETA SU	2.97e+00
21	53	55.2	2228	1	RRPL_SENDE	RNA POLYMERASE BETA SU	2.97e+00
22	52	54.2	109	1	PRVL_SALSA	PARVALBUMIN BETA 1 (CL	4.88e+00
23	52	54.2	188	1	RUVA_THEMEA	HOLLIDAY JUNCTION DNA	4.88e+00

24	52	54.2	732	1	CAD6_CHICK	B-CADHERIN PRECURSOR (4.88e+00
25	52	54.2	821	1	CAN3_RAT	CALPAIN P94, LARGE (CA	4.88e+00
26	52	54.2	821	1	CAN3_MOUSE	CALPAIN P94, LARGE (CA	4.88e+00
27	51	53.1	293	1	SPAO_SHFL	SURFACE PRESENTATION O	7.95e+00
28	51	53.1	644	1	YEAG_ECOLI	HYPOTHETICAL 74.5 KD P	7.95e+00
29	51	53.1	741	1	IDH2_VIBAI	ISOCITRATE DEHYDROGENA	7.95e+00
30	51	53.1	1783	1	Y468_MCGE	HYPOTHETICAL PROTEIN M	7.95e+00
31	50	52.1	217	1	IAP2_SOLTU	ASPARTIC PROTEINASE IN	1.29e+01
32	50	52.1	276	1	DHCA_RAT	CARBOXYL REDUCTASE (NA	1.29e+01
33	50	52.1	312	1	PRMA_STMAU	PROBABLE METHYLTRANSFER	1.29e+01
34	50	52.1	760	1	VACB_SHFL	VACB PROTEIN	1.29e+01
35	50	52.1	809	1	TER1_CAEEL	TRANSITIONAL ENDOPLASM	1.29e+01
36	50	52.1	2410	1	POLI_BAYM	GENOME POLYPROTEIN 1 (1.29e+01
37	50	52.1	2412	1	POLI_BAYM	GENOME POLYPROTEIN 1 (1.29e+01
38	49	51.0	266	1	ERMF_BACFR	RNA ADENINE N-6-METHY	2.06e+01
39	49	51.0	306	1	LUXD_PHOPO	ACYL TRANSFERASE (EC 2	2.06e+01
40	49	51.0	318	1	CYF_OENHO	APOCYTOCHROME F PRECUR	2.06e+01
41	49	51.0	320	1	CYF_TOBAC	APOCYTOCHROME F PRECUR	2.06e+01
42	49	51.0	320	1	CYF_VICPA	APOCYTOCHROME F PRECUR	2.06e+01
43	49	51.0	320	1	CYF_PEA	APOCYTOCHROME F PRECUR	2.06e+01
44	49	51.0	810	1	TER2_CAEEL	TRANSITIONAL ENDOPLASM	2.06e+01
45	49	51.0	916	1	CAD4_HUMAN	RETINAL-CADHERIN PRECU	2.06e+01

ALIGNMENTS

RESULT 1
 ID DSG3_HUMAN STANDARD: PRT; 999 AA.

AC P32926;
 DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).

GN DSG3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92069753.

RA Amagai M., Klaus-Kovtun V., Stanley J.R.;

RT "Autoantibodies against a novel epithelial cadherin in pemphigus

KT vulgaris, a disease of cell adhesion.";

RL Cell 67:869-877(1991).

CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND

CC CARCINOMAS.

CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS

CC (POTENTIAL).

CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN

CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE

CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES

CC AGAINST DSG3.

CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>

CC or send an email to license@sdb.ch).

CC EMBL: M76482; AAA60230.1; -;

DR PIR: A41088; IJH03.

DR HSSP: P09803; 1EDH.

DR MIM: 169615; -;

DR PROSITE: PS00232; CADHERIN: 3.

DR PFM: PF00028; cadherin: 4.

KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;

KW Calcium-binding: Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 30 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA; 107503 MW; 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 96; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 1.17e-11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ID DSG1_BOVIN STANDARD; PRT: 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 GN DSG1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RA MEDLINE: 91168965.
 RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
 RA Zimbelmann R., Franke W.W.;
 RT "Identification of desmoglein, a constitutive desmosomal
 RT glycoprotein, as a member of the cadherin family of cell adhesion
 RT molecules.";
 RL Eur. J. Cell Biol. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RA MEDLINE: 92037656.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
 RA Franke W.W.;
 RT "Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.";
 RL Eur. J. Cell Biol. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RC MEDLINE: 91097553.
 RA Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;
 RT "Desmoglein shows extensive homology to the cadherin family of cell
 RT adhesion molecules.";
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X58466; CAA41380.1; -;
 DR EMBL: X57784; CAA40930.1; -;
 DR EMBL: M58165; AAA62709.1; -;
 DR PIR: S14603; IYBOG1.
 DR HSSP: P09803; 1EDH.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PRAM: PF00028; cadherin; 3.
 DR Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 KW Calcium-binding: Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 DESMOGLEIN 1.
 FT CHAIN 50 1043 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 50 548 POTENTIAL.
 FT TRANSMEM 549 573 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 574 1043 CADHERIN 1.
 FT REPEAT 159 270 CADHERIN 2.
 FT REPEAT 271 385 CADHERIN 3.
 FT REPEAT 386 498 CADHERIN 4.
 FT REPEAT 419 845 DESMOGLEIN REPEAT 1.
 FT REPEAT 846 875 DESMOGLEIN REPEAT 2.
 FT REPEAT 876 905 DESMOGLEIN REPEAT 3.
 FT REPEAT 906 933 DESMOGLEIN REPEAT 4.
 FT REPEAT 934 962 DESMOGLEIN REPEAT 5.
 FT DOMAIN 963 1012 GLY/SER-RICH.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 496 496 POTENTIAL.
 FT CONFLICT 124 124 I -> V (IN REF. 4).
 SQ SEQUENCE 1043 AA; 112243 MW; ADE46133P8B77C11 CRC64;

Query Match 90.6%; Score 87; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 5.08e-09;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 Oy 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ID DSG1_HUMAN STANDARD; PRT: 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGI).
 GN DSG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RA Medley G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Arneemann J., Rutman A.J., Pidsley S.C., Watt F.M., Rees D.A.,
 RA Buxton R.S., Magee A.I.;
 RT "Desmosomal glycoprotein DGI, a component of intercellular desmosome

RT Junctions, is related to the cadherin family of cell adhesion
RT molecules."
RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56654; CAA39976.1; -
DR PIR: S16906; IJHUG1.
DR HSSP: P09803; IEDH.
DR MIM: 125670; -
DR PROSITE: PS00232; CADHERIN. 2.
DR PRAM: PF00028; cadherin. 4.
DR Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 23
FT PROPEP 24
FT CHAIN 49
FT DOMAIN 50 1049
FT TRANSSEM 545
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 813 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SQ SEQUENCE 1049 AA; 113715 MW; EEEL25655B9D6619 CRC64;
Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 5.08e-09;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 190 LNSKIAFKIVSOEPA 204
OY 1 LNSKIAFKIVSOEPA 15
RESULT 4
ID DSG2 HUMAN STANDARD; PRT: 1117 AA.
AC 014126;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 2 PRECURSOR (HDGC).
DS DSG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CARCINOMA;
RX MEDLINE: 94192736.

RA Schaefer S., Koch P.J., Franke W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the
RT expression catalogue of the desmoglein subfamily of desmosomal
RT cadherins."
RL Exp. Cell Res. 211:391-399(1994).
RN (2)
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE: 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R., Franke W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene."
RL Eur. J. Cell Biol. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26317; CAA81226.1; -
DR HSSP: P15116; INCI.
DR MIM: 125671; -
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PRAM: PF00028; cadherin. 4.
DR Cell adhesion; Signal; Transmembrane; Repeat; Signal;
KW Cytoskeleton; Calcium-binding.
FT SIGNAL 23
FT PROPEP 24 48
FT CHAIN 49 1117
FT DOMAIN 49 608
FT TRANSSEM 609 633
FT DOMAIN 634 1117
FT REPEAT 49 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 880 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SQ SEQUENCE 1117 AA; 122385 MW; 223B897FED70B289 CRC64;
Query Match 80.2%; Score 77; DB 1; Length 1117;
Best Local Similarity 73.3%; Pred. No. 3.24e-06;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 191 LNSKISRIYSLSPA 205
OY 1 LNSKIAFKIVSOEPA 15
RESULT 5
ID CAD1 MOUSE STANDARD; PRT: 884 AA.
AC P09803; 061377;
DT 01-MAR-1989 (Rel. 10, Created)

01-MAR-1989 (Rel. 10, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 EPIHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UVOGOLIN) (ARC-1).
 CDH1.
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 SEQUENCE FROM N.A.
 STRAIN-ICR.
 MEDLINE: 87315445.
 Nagafuchi A., Shirayoshi Y., Okazaki K., Yasuda K., Takeichi M.;
 "Transformation of cell adhesion properties by exogenously introduced
 E-cadherin cDNA.";
 Nature 329:341-343(1987).
 [2]
 SEQUENCE FROM N.A.
 STRAIN-129/SV.
 MEDLINE: 92093614.
 Ringwald M., Barbault H., Schmidt C., Kemler R.;
 "The structure of the gene coding for the mouse cell adhesion
 molecule uvomorulin.";
 Nucleic Acids Res. 19:6533-6539(1991).
 [3]
 SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.
 MEDLINE: 8811553.
 Ringwald M., Schuh R., Vestweber D., Eistetter H., Lottspeich F.,
 Engel J., Diehl R., Jaehnis F., Epplein J., Mayer S., Mueller C.,
 Kemler R.;
 "The structure of cell adhesion molecule uvomorulin. Insights into
 the molecular mechanism of Ca²⁺-dependent cell adhesion.";
 EMBO J. 6:3647-3653(1987).
 [4]
 SEQUENCE OF 1-15 FROM N.A.
 MEDLINE: 92107977.
 Behrens J., Loewerick O., Klein-Hitpass L., Birchmeier W.;
 "The E-cadherin promoter: functional analysis of a G.C-rich region
 and an epithelial cell-specific palindromic regulatory element.";
 Proc. Natl. Acad. Sci. U.S.A. 88:11495-11499(1991).
 [5]
 X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
 MEDLINE: 96176249.
 Negar B., Overduin M., Ikura M., Rini J.M.;
 "Structural basis of calcium-induced E-cadherin rigidification and
 dimerization.";
 Nature 380:360-364(1996).
 [6]
 STRUCTURE BY NMR OF 157-260.
 MEDLINE: 96271285.
 Overduin M., Tong K.I., Kay C.M., Ikura M.;
 "1H, 15N and 13C resonance assignments and monomeric structure of the
 amino-terminal extracellular domain of epithelial cadherin.";
 J. Biomol. NMR 7:173-189(1996).
 -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
 SORTING OF HETEROGENEOUS CELL TYPES.
 -!- SUBUNIT: HOMODIMER.
 -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 -!- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
 -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 or send an email to license@isb-sdb.ch).

 EMBL: X06115; CAA29488.1; -
 DR EMBL: X60961; CAA43292.1; -
 DR EMBL: X60962; CAA43292.1; JOINED.

DR EMBL: X60963; CAA43292.1; JOINED.
 DR EMBL: X60964; CAA43292.1; JOINED.
 DR EMBL: X60965; CAA43292.1; JOINED.
 DR EMBL: X60966; CAA43292.1; JOINED.
 DR EMBL: X60967; CAA43292.1; JOINED.
 DR EMBL: X60968; CAA43292.1; JOINED.
 DR EMBL: X60969; CAA43292.1; JOINED.
 DR EMBL: X60970; CAA43292.1; JOINED.
 DR EMBL: X60971; CAA43292.1; JOINED.
 DR EMBL: X60972; CAA43292.1; JOINED.
 DR EMBL: X60973; CAA43292.1; JOINED.
 DR EMBL: X60974; CAA43292.1; JOINED.
 DR EMBL: X60975; CAA43292.1; JOINED.
 DR EMBL: X60976; CAA43292.1; JOINED.
 DR EMBL: X60977; CAA43292.1; JOINED.
 DR EMBL: X60978; CAA43292.1; JOINED.
 DR EMBL: X60979; CAA43292.1; JOINED.
 DR EMBL: X60980; CAA43292.1; JOINED.
 DR EMBL: X60981; CAA43292.1; JOINED.
 DR EMBL: X60982; CAA43292.1; JOINED.
 DR EMBL: X60983; CAA43292.1; JOINED.
 DR EMBL: X60984; CAA43292.1; JOINED.
 DR EMBL: X60985; CAA43292.1; JOINED.
 DR EMBL: X60986; CAA43292.1; JOINED.
 DR EMBL: X60987; CAA43292.1; JOINED.
 DR EMBL: X60988; CAA43292.1; JOINED.
 DR EMBL: X60989; CAA43292.1; JOINED.
 DR EMBL: X60990; CAA43292.1; JOINED.
 DR EMBL: X60991; CAA43292.1; JOINED.
 DR EMBL: X60992; CAA43292.1; JOINED.
 DR EMBL: X60993; CAA43292.1; JOINED.
 DR EMBL: X60994; CAA43292.1; JOINED.
 DR EMBL: X60995; CAA43292.1; JOINED.
 DR EMBL: X60996; CAA43292.1; JOINED.
 DR EMBL: X60997; CAA43292.1; JOINED.
 DR EMBL: X60998; CAA43292.1; JOINED.
 DR EMBL: X60999; CAA43292.1; JOINED.
 DR EMBL: X61000; CAA43292.1; JOINED.
 DR EMBL: X61001; CAA43292.1; JOINED.
 DR EMBL: X61002; CAA43292.1; JOINED.
 DR EMBL: X61003; CAA43292.1; JOINED.
 DR EMBL: X61004; CAA43292.1; JOINED.
 DR EMBL: X61005; CAA43292.1; JOINED.
 DR EMBL: X61006; CAA43292.1; JOINED.
 DR EMBL: X61007; CAA43292.1; JOINED.
 DR EMBL: X61008; CAA43292.1; JOINED.
 DR EMBL: X61009; CAA43292.1; JOINED.
 DR EMBL: X61010; CAA43292.1; JOINED.
 DR EMBL: X61011; CAA43292.1; JOINED.
 DR EMBL: X61012; CAA43292.1; JOINED.
 DR EMBL: X61013; CAA43292.1; JOINED.
 DR EMBL: X61014; CAA43292.1; JOINED.
 DR EMBL: X61015; CAA43292.1; JOINED.
 DR EMBL: X61016; CAA43292.1; JOINED.
 DR EMBL: X61017; CAA43292.1; JOINED.
 DR EMBL: X61018; CAA43292.1; JOINED.
 DR EMBL: X61019; CAA43292.1; JOINED.
 DR EMBL: X61020; CAA43292.1; JOINED.
 DR EMBL: X61021; CAA43292.1; JOINED.
 DR EMBL: X61022; CAA43292.1; JOINED.
 DR EMBL: X61023; CAA43292.1; JOINED.
 DR EMBL: X61024; CAA43292.1; JOINED.
 DR EMBL: X61025; CAA43292.1; JOINED.
 DR EMBL: X61026; CAA43292.1; JOINED.
 DR EMBL: X61027; CAA43292.1; JOINED.
 DR EMBL: X61028; CAA43292.1; JOINED.
 DR EMBL: X61029; CAA43292.1; JOINED.
 DR EMBL: X61030; CAA43292.1; JOINED.
 DR EMBL: X61031; CAA43292.1; JOINED.
 DR EMBL: X61032; CAA43292.1; JOINED.
 DR EMBL: X61033; CAA43292.1; JOINED.
 DR EMBL: X61034; CAA43292.1; JOINED.
 DR EMBL: X61035; CAA43292.1; JOINED.
 DR EMBL: X61036; CAA43292.1; JOINED.
 DR EMBL: X61037; CAA43292.1; JOINED.
 DR EMBL: X61038; CAA43292.1; JOINED.
 DR EMBL: X61039; CAA43292.1; JOINED.
 DR EMBL: X61040; CAA43292.1; JOINED.
 DR EMBL: X61041; CAA43292.1; JOINED.
 DR EMBL: X61042; CAA43292.1; JOINED.
 DR EMBL: X61043; CAA43292.1; JOINED.
 DR EMBL: X61044; CAA43292.1; JOINED.
 DR EMBL: X61045; CAA43292.1; JOINED.
 DR EMBL: X61046; CAA43292.1; JOINED.
 DR EMBL: X61047; CAA43292.1; JOINED.
 DR EMBL: X61048; CAA43292.1; JOINED.
 DR EMBL: X61049; CAA43292.1; JOINED.
 DR EMBL: X61050; CAA43292.1; JOINED.
 DR EMBL: X61051; CAA43292.1; JOINED.
 DR EMBL: X61052; CAA43292.1; JOINED.
 DR EMBL: X61053; CAA43292.1; JOINED.
 DR EMBL: X61054; CAA43292.1; JOINED.
 DR EMBL: X61055; CAA43292.1; JOINED.
 DR EMBL: X61056; CAA43292.1; JOINED.
 DR EMBL: X61057; CAA43292.1; JOINED.
 DR EMBL: X61058; CAA43292.1; JOINED.
 DR EMBL: X61059; CAA43292.1; JOINED.
 DR EMBL: X61060; CAA43292.1; JOINED.
 DR EMBL: X61061; CAA43292.1; JOINED.
 DR EMBL: X61062; CAA43292.1; JOINED.
 DR EMBL: X61063; CAA43292.1; JOINED.
 DR EMBL: X61064; CAA43292.1; JOINED.
 DR EMBL: X61065; CAA43292.1; JOINED.
 DR EMBL: X61066; CAA43292.1; JOINED.
 DR EMBL: X61067; CAA43292.1; JOINED.
 DR EMBL: X61068; CAA43292.1; JOINED.
 DR EMBL: X61069; CAA43292.1; JOINED.
 DR EMBL: X61070; CAA43292.1; JOINED.
 DR EMBL: X61071; CAA43292.1; JOINED.
 DR EMBL: X61072; CAA43292.1; JOINED.
 DR EMBL: X61073; CAA43292.1; JOINED.
 DR EMBL: X61074; CAA43292.1; JOINED.
 DR EMBL: X61075; CAA43292.1; JOINED.
 DR EMBL: X61076; CAA43292.1; JOINED.
 DR EMBL: X61077; CAA43292.1; JOINED.
 DR EMBL: X61078; CAA43292.1; JOINED.
 DR EMBL: X61079; CAA43292.1; JOINED.
 DR EMBL: X61080; CAA43292.1; JOINED.
 DR EMBL: X61081; CAA43292.1; JOINED.
 DR EMBL: X61082; CAA43292.1; JOINED.
 DR EMBL: X61083; CAA43292.1; JOINED.
 DR EMBL: X61084; CAA43292.1; JOINED.
 DR EMBL: X61085; CAA43292.1; JOINED.
 DR EMBL: X61086; CAA43292.1; JOINED.
 DR EMBL: X61087; CAA43292.1; JOINED.
 DR EMBL: X61088; CAA43292.1; JOINED.
 DR EMBL: X61089; CAA43292.1; JOINED.
 DR EMBL: X61090; CAA43292.1; JOINED.
 DR EMBL: X61091; CAA43292.1; JOINED.
 DR EMBL: X61092; CAA43292.1; JOINED.
 DR EMBL: X61093; CAA43292.1; JOINED.
 DR EMBL: X61094; CAA43292.1; JOINED.
 DR EMBL: X61095; CAA43292.1; JOINED.
 DR EMBL: X61096; CAA43292.1; JOINED.
 DR EMBL: X61097; CAA43292.1; JOINED.
 DR EMBL: X61098; CAA43292.1; JOINED.
 DR EMBL: X61099; CAA43292.1; JOINED.
 DR EMBL: X61100; CAA43292.1; JOINED.
 DR EMBL: X61101; CAA43292.1; JOINED.
 DR EMBL: X61102; CAA43292.1; JOINED.
 DR EMBL: X61103; CAA43292.1; JOINED.
 DR EMBL: X61104; CAA43292.1; JOINED.
 DR EMBL: X61105; CAA43292.1; JOINED.
 DR EMBL: X61106; CAA43292.1; JOINED.
 DR EMBL: X61107; CAA43292.1; JOINED.
 DR EMBL: X61108; CAA43292.1; JOINED.
 DR EMBL: X61109; CAA43292.1; JOINED.
 DR EMBL: X61110; CAA43292.1; JOINED.
 DR EMBL: X61111; CAA43292.1; JOINED.
 DR EMBL: X61112; CAA43292.1; JOINED.
 DR EMBL: X61113; CAA43292.1; JOINED.
 DR EMBL: X61114; CAA43292.1; JOINED.
 DR EMBL: X61115; CAA43292.1; JOINED.
 DR EMBL: X61116; CAA43292.1; JOINED.
 DR EMBL: X61117; CAA43292.1; JOINED.
 DR EMBL: X61118; CAA43292.1; JOINED.
 DR EMBL: X61119; CAA43292.1; JOINED.
 DR EMBL: X61120; CAA43292.1; JOINED.
 DR EMBL: X61121; CAA43292.1; JOINED.
 DR EMBL: X61122; CAA43292.1; JOINED.
 DR EMBL: X61123; CAA43292.1; JOINED.
 DR EMBL: X61124; CAA43292.1; JOINED.
 DR EMBL: X61125; CAA43292.1; JOINED.
 DR EMBL: X61126; CAA43292.1; JOINED.
 DR EMBL: X61127; CAA43292.1; JOINED.
 DR EMBL: X61128; CAA43292.1; JOINED.
 DR EMBL: X61129; CAA43292.1; JOINED.
 DR EMBL: X61130; CAA43292.1; JOINED.
 DR EMBL: X61131; CAA43292.1; JOINED.
 DR EMBL: X61132; CAA43292.1; JOINED.
 DR EMBL: X61133; CAA43292.1; JOINED.
 DR EMBL: X61134; CAA43292.1; JOINED.
 DR EMBL: X61135; CAA43292.1; JOINED.
 DR EMBL: X61136; CAA43292.1; JOINED.
 DR EMBL: X61137; CAA43292.1; JOINED.
 DR EMBL: X61138; CAA43292.1; JOINED.
 DR EMBL: X61139; CAA43292.1; JOINED.
 DR EMBL: X61140; CAA43292.1; JOINED.
 DR EMBL: X61141; CAA43292.1; JOINED.
 DR EMBL: X61142; CAA43292.1; JOINED.
 DR EMBL: X61143; CAA43292.1; JOINED.
 DR EMBL: X61144; CAA43292.1; JOINED.
 DR EMBL: X61145; CAA43292.1; JOINED.
 DR EMBL: X61146; CAA43292.1; JOINED.
 DR EMBL: X61147; CAA43292.1; JOINED.
 DR EMBL: X61148; CAA43292.1; JOINED.
 DR EMBL: X61149; CAA43292.1; JOINED.
 DR EMBL: X61150; CAA43292.1; JOINED.
 DR EMBL: X61151; CAA43292.1; JOINED.
 DR EMBL: X61152; CAA43292.1; JOINED.
 DR EMBL: X61153; CAA43292.1; JOINED.
 DR EMBL: X61154; CAA43292.1; JOINED.
 DR EMBL: X61155; CAA43292.1; JOINED.
 DR EMBL: X61156; CAA43292.1; JOINED.
 DR EMBL: X61157; CAA43292.1; JOINED.
 DR EMBL: X61158; CAA43292.1; JOINED.
 DR EMBL: X61159; CAA43292.1; JOINED.
 DR EMBL: X61160; CAA43292.1; JOINED.
 DR EMBL: X61161; CAA43292.1; JOINED.
 DR EMBL: X61162; CAA43292.1; JOINED.
 DR EMBL: X61163; CAA43292.1; JOINED.
 DR EMBL: X61164; CAA43292.1; JOINED.
 DR EMBL: X61165; CAA43292.1; JOINED.
 DR EMBL: X61166; CAA43292.1; JOINED.
 DR EMBL: X61167; CAA43292.1; JOINED.
 DR EMBL: X61168; CAA43292.1; JOINED.
 DR EMBL: X61169; CAA43292.1; JOINED.
 DR EMBL: X61170; CAA43292.1; JOINED.
 DR EMBL: X61171; CAA43292.1; JOINED.
 DR EMBL: X61172; CAA43292.1; JOINED.
 DR EMBL: X61173; CAA43292.1; JOINED.
 DR EMBL: X61174; CAA43292.1; JOINED.
 DR EMBL: X61175; CAA43292.1; JOINED.
 DR EMBL: X61176; CAA43292.1; JOINED.
 DR EMBL: X61177; CAA43292.1; JOINED.
 DR EMBL: X61178; CAA43292.1; JOINED.
 DR EMBL: X61179; CAA43292.1; JOINED.
 DR EMBL: X61180; CAA43292.1; JOINED.
 DR EMBL: X61181; CAA43292.1; JOINED.
 DR EMBL: X61182; CAA43292.1; JOINED.
 DR EMBL: X61183; CAA43292.1; JOINED.
 DR EMBL: X61184; CAA43292.1; JOINED.
 DR EMBL: X61185; CAA43292.1; JOINED.
 DR EMBL: X61186; CAA43292.1; JOINED.
 DR EMBL: X61187; CAA43292.1; JOINED.
 DR EMBL: X61188; CAA43292.1; JOINED.
 DR EMBL: X61189; CAA43292.1; JOINED.
 DR EMBL: X61190; CAA43292.1; JOINED.
 DR EMBL: X61191; CAA43292.1; JOINED.
 DR EMBL: X61192; CAA43292.1; JOINED.
 DR EMBL: X61193; CAA43292.1; JOINED.
 DR EMBL: X61194; CAA43292.1; JOINED.
 DR EMBL: X61195; CAA43292.1; JOINED.
 DR EMBL: X61196; CAA43292.1; JOINED.
 DR EMBL: X61197; CAA43292.1; JOINED.
 DR EMBL: X61198; CAA43292.1; JOINED.
 DR EMBL: X61199; CAA43292.1; JOINED.
 DR EMBL: X61200; CAA43292.1; JOINED.
 DR EMBL: X61201; CAA43292.1; JOINED.
 DR EMBL: X61202; CAA43292.1; JOINED.
 DR EMBL: X61203; CAA43292.1; JOINED.
 DR EMBL: X61204; CAA43292.1; JOINED.
 DR EMBL: X61205; CAA43292.1; JOINED.
 DR EMBL: X61206; CAA43292.1; JOINED.
 DR EMBL: X61207; CAA43292.1; JOINED.
 DR EMBL: X61208; CAA43292.1; JOINED.
 DR EMBL: X61209; CAA43292.1; JOINED.
 DR EMBL: X61210; CAA43292.1; JOINED.
 DR EMBL: X61211; CAA43292.1; JOINED.
 DR EMBL: X61212; CAA43292.1; JOINED.
 DR EMBL: X61213; CAA43292.1; JOINED.
 DR EMBL: X61214; CAA43292.1; JOINED.
 DR EMBL: X61215; CAA43292.1; JOINED.
 DR EMBL: X61216; CAA43292.1; JOINED.
 DR EMBL: X61217; CAA43292.1; JOINED.
 DR EMBL: X61218; CAA43292.1; JOINED.
 DR EMBL: X61219; CAA43292.1; JOINED.
 DR EMBL: X61220; CAA43292.1; JOINED.
 DR EMBL: X61221; CAA43292.1; JOINED.
 DR EMBL: X61222; CAA43292.1; JOINED.
 DR EMBL: X61223; CAA43292.1; JOINED.
 DR EMBL: X61224; CAA43292.1; JOINED.
 DR EMBL: X61225; CAA43292.1; JOINED.
 DR EMBL: X61226; CAA43292.1; JOINED.
 DR EMBL: X61227; CAA43292.1; JOINED.
 DR EMBL: X61228; CAA43292.1; JOINED.
 DR EMBL: X61229; CAA43292.1; JOINED.
 DR EMBL: X61230; CAA43292.1; JOINED.
 DR EMBL: X61231; CAA43292.1; JOINED.
 DR EMBL: X61232; CAA43292.1; JOINED.
 DR EMBL: X61233; CAA43292.1; JOINED.
 DR EMBL: X61234; CAA43292.1; JOINED.
 DR EMBL: X61235; CAA43292.1; JOINED.
 DR EMBL: X61236; CAA43292.1; JOINED.
 DR EMBL: X61237; CAA43292.1; JOINED.
 DR EMBL: X61238; CAA43292.1; JOINED.
 DR EMBL: X61239; CAA43292.1; JOINED.
 DR EMBL: X61240; CAA43292.1; JOINED.
 DR EMBL: X61241; CAA43292.1; JOINED.
 DR EMBL: X61242; CAA43292.1; JOINED.
 DR EMBL: X61243; CAA43292.1; JOINED.
 DR EMBL: X61244; CAA43292.1; JOINED.
 DR EMBL: X61245; CAA43292.1; JOINED.
 DR EMBL: X61246; CAA43292.1; JOINED.
 DR EMBL: X61247; CAA43292.1; JOINED.
 DR EMBL: X61248; CAA43292.1; JOINED.
 DR EMBL: X61249; CAA43292.1; JOINED.
 DR EMBL: X61250; CAA43292.1; JOINED.
 DR EMBL: X61251; CAA43292.1; JOINED.
 DR EMBL: X61252; CAA43292.1; JOINED.
 DR EMBL: X61253; CAA43292.1; JOINED.
 DR EMBL: X61254; CAA43292.1; JOINED.
 DR EMBL: X61255; CAA43292.1; JOINED.
 DR EMBL: X61256; CAA43292.1; JOINED.
 DR EMBL: X61257; CAA43292.1; JOINED.
 DR EMBL: X61258; CAA43292.1; JOINED.
 DR EMBL: X61259; CAA43292.1; JOINED.
 DR EMBL: X61260; CAA43292.1; JOINED.
 DR EMBL: X61261; CAA43292.1; JOINED.
 DR EMBL: X61262; CAA43292.1; JOINED.
 DR EMBL: X61263; CAA43292.1; JOINED.
 DR EMBL: X61264; CAA43292.1; JOINED.
 DR EMBL: X61265; CAA43292.1; JOINED.
 DR EMBL: X61266; CAA43292.1; JOINED.
 DR EMBL: X61267; CAA43292.1; JOINED.
 DR EMBL: X61268; CAA43292.1; JOINED.
 DR EMBL: X61269; CAA43292.1; JOINED.
 DR EMBL: X61270; CAA43292.1; JOINED.
 DR EMBL: X61271; CAA43292.1; JOINED.
 DR EMBL: X61272; CAA43292.1; JOINED.
 DR EMBL: X61273; CAA43292.1; JOINED.
 DR EMBL: X61274; CAA43292.1; JOINED.
 DR EMBL: X61275; CAA43292.1; JOINED.
 DR EMBL: X61276; CAA43292.1; JOINED.
 DR EMBL: X61277; CAA43292.1; JOINED.
 DR EMBL: X61278; CAA43292.1; JOINED.
 DR EMBL: X61279; CAA43292.1; JOINED.
 DR EMBL: X61280; CAA43292.1; JOINED.
 DR EMBL: X61281; CAA43292.1; JOINED.
 DR EMBL: X61282; CAA43292.1; JOINED.
 DR EMBL: X61283; CAA43292.1; JOINED.
 DR EMBL: X61284; CAA43292.1; JOINED.
 DR EMBL: X61285; CAA43292.1; JOINED.
 DR EMBL: X61286; CAA43292.1; JOINED.
 DR EMBL: X61287; CAA43292.1; JOINED.
 DR EMBL: X61288; CAA43292.1; JOINED.
 DR EMBL: X61289; CAA43292.1; JOINED.
 DR EMBL: X61290; CAA43292.1; JOINED.
 DR EMBL: X61291; CAA43292.1; JOINED.
 DR EMBL: X61292; CAA43292.1; JOINED.
 DR EMBL: X61293; CAA43292.1; JOINED.
 DR EMBL: X61294; CAA43292.1; JOINED.
 DR EMBL: X61295; CAA43292.1; JOINED.
 DR EMBL: X61296; CAA43292.1; JOINED.
 DR EMBL: X61297; CAA43292.1; JOINED.
 DR EMBL: X61298; CAA43292.1; JOINED.
 DR EMBL: X61299; CAA43292.1; JOINED.
 DR EMBL: X61300; CAA43292.1; JOINED.
 DR EMBL: X61301; CAA43292.1; JOINED.
 DR EMBL: X61302; CAA43292.1; JOINED.
 DR EMBL: X61303; CAA43292.1; JOINED.
 DR EMBL: X61304; CAA43292.1; JOINED.
 DR EMBL: X61305; CAA43292.1; JOINED.
 DR EMBL: X61306; CAA43292.1; JOINED.
 DR EMBL: X61307; CAA43292.1; JOINED.
 DR EMBL: X61308; CAA43292.1; JOINED.
 DR EMBL: X61309; CAA43292.1; JOINED.
 DR EMBL: X61310; CAA43292.1; JOINED.
 DR EMBL: X61311; CAA43292.1; JOINED.
 DR EMBL: X61312; CAA43292.1; JOINED.
 DR EMBL: X61313; CAA43292.1; JOINED.
 DR EMBL: X61314; CAA43292.1; JOINED.
 DR EMBL: X61315; CAA43292.1; JOINED.
 DR EMBL: X61316; CAA43292.1; JOINED.
 DR EMBL: X61317; CAA43292.1; JOINED.
 DR EMBL: X61318; CAA43292.1; JOINED.
 DR EMBL: X61319; CAA43292.1; JOINED.
 DR EMBL: X61320; CAA43292.1; JOINED.
 DR EMBL: X61321; CAA43292.1; JOINED.
 DR EMBL: X61322; CAA43292.1; JOINED.
 DR EMBL: X61323; CAA43292.1; JOINED.
 DR EMBL: X61324; CAA43292.1; JOINED.
 DR EMBL: X61325; CAA43292.1; JOINED.
 DR EMBL: X61326; CAA43292.1; JOINED.
 DR EMBL: X61327; CAA43292.1; JOINED.
 DR EMBL: X61328; CAA43292.1; JOINED.
 DR EMBL: X61329; CAA43292.1; JOINED.
 DR EMBL: X61330; CAA43292.1; JOINED.
 DR EMBL: X61331; CAA43292.1; JOINED.
 DR EMBL: X61332; CAA43292.1; JOINED.
 DR EMBL: X61333; CAA43292.1

```

RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE: 92363956.
RA Salomon D., Ayala O., Patel-King R., Hynes R.O., Geiger B.:
RT "Extrajunctional distribution of N-cadherin in cultured human
RT endothelial cells."
RL J. Cell Sci. 102:7-17(1992).
RN [4]
RN SEQUENCE OF 160-906 FROM N.A.
RX MEDLINE: 90347462.
RA Walsh F.S., Barton C.H., Putt W., Moore S.E., Kelsell D.,
RA Spurr N., Goodfellow P.N.:
RT "N-cadherin gene maps to human chromosome 18 and is not linked to the
RT E-cadherin gene."
RL J. Neurochem. 55:805-812(1990).
RN [5]
RN SEQUENCE OF 1-20 FROM N.A.
RX MEDLINE: 95048366.
RA Wallis J.A., Fox M., Walsh F.S.:
RT "Structure of the human N-cadherin gene: VAC analysis and fine
RT chromosomal mapping to 18q11.2."
RL Genomics 22:172-179(1994).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57548; CAA40773.1; -
DR EMBL: X54315; CAA38213.1; -
DR EMBL: S43303; AAB22854.1; -
DR EMBL: M34064; AAB03336.1; -
DR EMBL: Z27420; CAA81799.1; -
DR PIR: A38870; IJHUCN.
DR HSSP: P15116; INCH.
DR MIM: 114020.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 3.
DR PIR: PF01049; Cadherin_C-term; 1.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat; Signal.
FT SIGNAL 1 23
FT PROPEP 24 159
FT CHAIN 160 906
FT DOMAIN 160 724
FT TRANSMEM 725 746
FT DOMAIN 747 906
FT REPEAT 160 267
FT REPEAT 268 382
FT REPEAT 383 497
FT REPEAT 498 603
FT REPEAT 604 714
FT DOMAIN 714 878
FT CARBOHYD 190 190
FT CARBOHYD 273 273
FT CARBOHYD 325 325
FT CARBOHYD 402 402
FT CARBOHYD 572 572
FT CARBOHYD 651 651
FT CARBOHYD 692 692
FT CONFLICT 12 12
FT CONFLICT 16 16
L -> A (IN REF. 3 AND 5).

```

```

FT CONFLICT 196 196 S -> T (IN REF. 1).
FT CONFLICT 212 212 I -> L (IN REF. 4).
FT CONFLICT 357 357 N -> I (IN REF. 1).
FT CONFLICT 867 867 A -> L (IN REF. 3).
SQ SEQUENCE 906 AA; 99851 MW; 72DDC7BB857C7AFC CRC64;

Query Match 62.5%; Score 60; DB 1; Length 906;
Best Local Similarity 46.7%; Pred. No. 7.53e-02;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB 300 LNSKIAFKIVSOEPA 15
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 7
ID YOX1_YEAST STANDARD: PRT; 385 AA.
AC P34161;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOMEOBOX PROTEIN YOX1.
GN YOX1 OR YML027W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RN SEQUENCE OF 1-296 FROM N.A.
RX MEDLINE: 93209080.
RA Kaufmann E.;
RT "In vitro binding to the leucine tRNA gene identifies a novel yeast
RT homeobox gene."
RL Chromosoma 102:174-179(1993).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO, IS CAPABLE OF BINDING TO THE DNA OF THE
CC LEUCINE tRNA GENE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62392; CAA44264.1; -
DR EMBL: Z46659; CAA86628.1; -
DR PIR: S33388; S33388.
DR HSSP: P06601; IFJL.
DR SGD: L0002540; YOX1.
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
DR PIR: PF00046; homeobox; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT DNA_BIND 176 235
FT CONFLICT 291 296
FT CONFLICT 291 296
SQ SEQUENCE 385 AA; 42739 MW; F21A95CDD97A282 CRC64;

Query Match 61.5%; Score 59; DB 1; Length 385;
Best Local Similarity 63.6%; Pred. No. 1.30e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 364 KFEKVIDQOP 374
Oy 4 KFEKVIDQOP 14

```

```

RESULT      8
ID YGN1_YEAST STANDARD: PRT: 1264 AA.
AC P53125:
DC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 145.6 kD PROTEIN IN RPLB-CG1 INTERGENIC REGION.
GN YGL13W OR G2842.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RA MEDLINE: 96437978.
RX Escribano V., Erasó P., Portillo F., Mazon M.J.;
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
RT cerevisiae chromosome VII reveals SEC27, SSM1b, a putative
RT S-adenosylmethionine-dependent enzyme and six new open reading
RT frames."
RL Yeast 12:887-892(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 272655; CAA96844.1;
DR EMBL: 272655; CAA96844.1;
DR HYPOTHETICAL protein.
SQ SEQUENCE 1264 AA: 145642 MW: 45E4CF8835C7C746 CRC64;
Query Match 59.48; Score 57; DB 1; Length 1264;
Jest Local Similarity 42.9%; Pred. No. 3,79e-01;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Db 524 SDKVEKIVDDPA 537
QY 2 NSKIARIVSOEPA 15
RESULT      9
ID FAT_DROME STANDARD: PRT: 5147 AA.
AC P33450:
DC 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN FT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92069752.
RA Mahoney P.A., Weber U., Onofrechuk P., Blessmann H., Bryant P.J.,
RA Goodman C.S.;
RT "The fat tumor suppressor gene in Drosophila encodes a novel member
RT of the cadherin gene superfamily."
RL Cell 67:853-868(1991).
CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC DIFFERENTIATION AND MORPHOGENESIS, AND DEATH DURING THE PUPAL
CC STAGE.
CC -1- MISCELLANEOUS: SIMILARITY: CONTAINS 37 CADHERINS-TYPE REPEATS.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M80537; AAA28530.1;
DR PIR: A41087; IJFPTM.
DR HSSP: P00740; IIXA.
DR ELYBASE: FBG0001075; ft.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN. 22.
DR PROSITE: PS00022; EGF_1; 4.
DR PROSITE: PS01186; EGF_2; 2.
DR PFM: PF00008; EGF; 4.
DR PFM: PF00028; cadherin; 34.
DR PFM: PF00054; laminin_G; 2.
KM Cell adhesion. Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KM Calcium-binding; Repeat; EGF-like domain.
KW SIGNAL 1 35
KW CHAIN 36 5147
KW DOMAIN 36 4583
KW TRANSMEM 4584 4609
KW DOMAIN 4610 5147
FT REPEAT 36 156
FT REPEAT 157 270
FT REPEAT 271 382
FT REPEAT 383 494
FT REPEAT 495 599
FT REPEAT 600 708
FT REPEAT 709 821
FT REPEAT 821 942
FT REPEAT 943 1049
FT REPEAT 1050 1153
FT REPEAT 1154 1278
FT REPEAT 1279 1384
FT REPEAT 1385 1489
FT REPEAT 1490 1601
FT REPEAT 1602 1713
FT REPEAT 1714 1823
FT REPEAT 1824 1922
FT REPEAT 1923 2027
FT REPEAT 2028 2167
FT REPEAT 2168 2278
FT REPEAT 2279 2385
FT REPEAT 2386 2491
FT REPEAT 2492 2596
FT REPEAT 2597 2703
FT REPEAT 2704 2810
FT REPEAT 2811 2913
FT REPEAT 2914 3013
FT REPEAT 3014 3124
FT REPEAT 3125 3229
FT REPEAT 3230 3334
FT REPEAT 3335 3439
FT REPEAT 3440 3545
FT REPEAT 3546 3651
FT REPEAT 3652 3756
FT REPEAT 3757 3854
FT REPEAT 3855 3959
FT REPEAT 3960 4011
FT REPEAT 4012 4049
FT REPEAT 4050 4090
FT REPEAT 4091 4128
FT REPEAT 4129 4231
FT REPEAT 4232 4362
FT REPEAT 4363 3966
FT REPEAT 4364 3999
FT REPEAT 4365 4010
FT REPEAT 4366 4028
FT REPEAT 4367 4037
FT REPEAT 4368 4048
FT REPEAT 4369 4067
FT DISULFID 4056

```

```

FT DISULFID 4061 4078 BY SIMILARITY.
FT DISULFID 4080 4089 BY SIMILARITY.
FT DISULFID 4096 4107 BY SIMILARITY.
FT DISULFID 4101 4116 BY SIMILARITY.
FT DISULFID 4118 4127 BY SIMILARITY.
FT DISULFID 4325 4341 BY SIMILARITY.
FT DISULFID 4334 4350 BY SIMILARITY.
FT DISULFID 4332 4351 BY SIMILARITY.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 280 280 POTENTIAL.
FT CARBOHYD 402 402 POTENTIAL.
FT CARBOHYD 461 461 POTENTIAL.
FT CARBOHYD 605 605 POTENTIAL.
FT CARBOHYD 631 631 POTENTIAL.
FT CARBOHYD 1155 1155 POTENTIAL.
FT CARBOHYD 1367 1367 POTENTIAL.
FT CARBOHYD 1458 1458 POTENTIAL.
FT CARBOHYD 1751 1751 POTENTIAL.
FT CARBOHYD 1831 1831 POTENTIAL.
FT CARBOHYD 1880 1880 POTENTIAL.
FT CARBOHYD 2080 2080 POTENTIAL.
FT CARBOHYD 2171 2171 POTENTIAL.
FT CARBOHYD 2247 2247 POTENTIAL.
FT CARBOHYD 2290 2290 POTENTIAL.
FT CARBOHYD 2437 2437 POTENTIAL.
FT CARBOHYD 2581 2581 POTENTIAL.
FT CARBOHYD 2799 2799 POTENTIAL.
FT CARBOHYD 2920 2920 POTENTIAL.
FT CARBOHYD 2946 2946 POTENTIAL.
FT CARBOHYD 2967 2967 POTENTIAL.
FT CARBOHYD 3157 3157 POTENTIAL.
FT CARBOHYD 3303 3303 POTENTIAL.
FT CARBOHYD 3386 3386 POTENTIAL.
FT CARBOHYD 3389 3389 POTENTIAL.
FT CARBOHYD 3525 3525 POTENTIAL.
FT CARBOHYD 3852 3852 POTENTIAL.
FT CARBOHYD 3865 3865 POTENTIAL.
FT CARBOHYD 3905 3905 POTENTIAL.
FT CARBOHYD 4306 4306 POTENTIAL.
FT CARBOHYD 4414 4414 POTENTIAL.
FT CARBOHYD 4471 4471 POTENTIAL.
FT CARBOHYD 4487 4487 POTENTIAL.
FT CARBOHYD 4539 4539 POTENTIAL.
FT CARBOHYD 4550 4550 POTENTIAL.
FT VARIANT 1229 1229 S -> G.
FT VARIANT 1233 1233 G -> S.
SQ SEQUENCE 5147 AA; 564868 MW; 4BE85ED6C12397EE CRC64;

Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 3.79e-01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 1520 LNSKIATKIVSOEP 1533
Oy 1 LNSKIATKIVSOEP 14

RESULT 10
ID EFP_BORBU STANDARD: PRT; 192 AA.
AC 051232;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE ELONGATION FACTOR P (EF-P).
GN EFP OR BB0214.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC 35210 / B31;
RX MEDLINE: 98065943.

```

```

RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uitterback T., Matthey L., McDonald L., Atlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
CC -1- FUNCTION: INVOLVED IN PEPTIDE BOND SYNTHESIS. STIMULATE EFFICIENT
CC TRANSLATION AND PEPTIDE-BOND SYNTHESIS ON NATIVE OR RECONSTITUTED
CC 70S RIBOSOMES IN VITRO. PROBABLY FUNCTIONS INDIRECTLY BY ALTERING
CC THE AFFINITY OF THE RIBOSOME FOR AMINOACYL-TRNA, THUS INCREASING
CC THEIR REACTIVITY AS ACCEPTORS FOR PEPTIDYL TRANSFERASE
CC (BY SIMILARITY).
CC -1- PATHWAY: PROTEIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ELONGATION FACTOR P FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC .....
CC DR EMBL; AE001132; AAC66610.1; -.
CC TIGR; BB0214; -.
DR PROSITE; PS01275; EFP; FALSE_NEG.
DR PFAM; PF01132; EFP; 1.
KW Protein biosynthesis; Elongation factor.
SQ SEQUENCE 192 AA; 21413 MW; AC2C48AB8C73DE0 CRC64;

```

```

Query Match 57.3%; Score 55; DB 1; Length 192;
Best Local Similarity 53.3%; Pred. 1.08e-00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

Db 130 LARKIEVEVEVEA 144
Oy 1 LNSKIATKIVSOEP 15

RESULT 11
ID CAD2_BOVIN STANDARD: PRT; 877 AA.
AC P19534;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NEURAL-CADHERIN PRECURSOR (N-CADHERIN) (FRAGMENT).
GN CDR2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90360979.
RA Liaw C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RT endothelial cells".
RL EMBL J. 9:2701-2708(1990).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).

DR EMBL: X53615: CA37677.1: -
 DR PIR: S11693: IBOCN.
 DR HSP: S11693: INCH.
 DR PROSITE: PS00232: CADHERIN: 3.
 DR PFAM: PF00028: cadherin: 5.
 DR PFAM: PF01049: Cadherin_C-term: 1.
 KM Cell adhesion: Glycoprotein: Phosphorylation: Transmembrane:
 KW Calcium-binding: Repeat.
 FT NON_TER 1
 FT PROPEP <1 130
 FT CHAIN 131 877 NEURAL-CADHERIN.
 FT DOMAIN 131 695 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 696 717 POTENTIAL.
 FT DOMAIN 718 877 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 131 238 CADHERIN 1.
 FT REPEAT 239 363 CADHERIN 2.
 FT REPEAT 364 468 CADHERIN 3.
 FT REPEAT 469 574 CADHERIN 4.
 FT REPEAT 575 685 CADHERIN 5.
 FT DOMAIN 834 849 SER-RICH.
 FT CARBOHYD 161 161 POTENTIAL.
 FT CARBOHYD 244 244 POTENTIAL.
 FT CARBOHYD 296 296 POTENTIAL.
 FT CARBOHYD 373 373 POTENTIAL.
 FT CARBOHYD 543 543 POTENTIAL.
 FT CARBOHYD 622 622 POTENTIAL.
 FT CARBOHYD 663 663 POTENTIAL.
 SO SEQUENCE 877 AA: 96845 MW: 44188295ED871A249 CRC64:
 Query Match 56.3% Score 54; DB 1; Length 877;
 Best Local Similarity 40.0%; Pred. No. 1.79e+00;
 Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 271 LNCMRITLSOAPS 285
 1 LNSKIAFKIVSOEPA 15

RESULT 12
 ID CADI_HUMAN STANDARD: PRT: 882 AA.
 AC P12830: Q14216:
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UWOMORULIN) (CAM 120/80).
 GN CDH1 OR UVO OR CDHE.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93211394.
 RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debryne F.M.J.,
 RA Schalken J.A.;
 RT "Molecular cloning and characterization of the human E-cadherin
 RT CDNA.";
 RL Mol. Biol. Rep. 17:123-128(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kelder W., Warda A., Oda T., Hirohashi S., Kemler R., Birchmeier W.;
 RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 172-311 FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 89031725.
 RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.;
 RT "Characterization and chromosomal localization of the gene encoding

RT the human cell adhesion molecule uvomorulin.";
 RL Differentiation 38:67-71(1988).
 RN [4]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE: 94380041.
 RA Bussemakers M.J., Giroldi L.A., van Bokhoven A., Schalken J.A.;
 RT "Transcriptional regulation of the human E-cadherin gene in human
 RT prostate cancer cell lines: Characterization of the human E-cadherin
 RT gene promoter.";
 RL Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
 RN [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 98415721.
 RA Berr G., Becker K.-F., Hoefler H., van Roy F.;
 RT "Mutations of the human E-cadherin (CDH1) gene.";
 RL Hum. Mutat. 12:226-237(1998).
 RN [6]
 RP VARIANTS ALA-370 AND ASN-473.
 RX MEDLINE: 94306394.
 RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
 RT "E-cadherin gene mutations provide clues to diffuse type gastric
 RT carcinomas.";
 RL Cancer Res. 54:3845-3852(1994).
 RN [7]
 RP VARIANT LOBULAR BREAST CARCINOMA SER-315.
 RX MEDLINE: 95049851.
 RA Kanal Y., Oda T., Tsuda H., Ochiai A., Hirohashi S.;
 RT "Point mutation of the E-cadherin gene in invasive lobular carcinoma
 RT of the breast.";
 RL Jpn. J. Cancer Res. 85:1035-1039(1994).
 RN [8]
 RP VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
 RX MEDLINE: 94353985.
 RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.;
 RT "Mutations of the E-cadherin gene in human gynecologic cancers.";
 RL Nat. Genet. 7:98-102(1994).
 RN [9]
 RP VARIANT GASTRIC ADENOCARCINOMA GLY-274--PRO-277 DEL.
 RX MEDLINE: 94173928.
 RA Oda T., Kanal Y., Oyama T., Yoshitura K., Shimoyama Y., Birchmeier W.,
 RA Sugimura T., Hirohashi S.;
 RT "E-cadherin gene mutations in human gastric carcinoma cell lines.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
 RN [10]
 RP VARIANT GASTRIC CARCINOMA PRO-193.
 RX MEDLINE: 96390918.
 RA Muta H., Noguchi M., Kanal Y., Ochiai A., Nawata H., Hirohashi S.;
 RT "E-cadherin gene mutations in signet ring cell carcinoma of the
 RT stomach.";
 RL Jpn. J. Cancer Res. 87:843-848(1996).
 RN [11]
 RP VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND ASP-418--PHE-423 DEL.
 RX MEDLINE: 97197648.
 RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
 RA Terashima M., Saito K., Satodate R.;
 RT "Inactivation of the E-cadherin gene in primary gastric carcinomas and
 RT gastric carcinoma cell lines.";
 RL Jpn. J. Cancer Res. 87:1153-1159(1996).
 RN [12]
 RP VARIANT THYROID TUMOR THR-592.
 RX MEDLINE: 97138061.
 RA Soares P., Berr G., van Roy F., Sobrinho-Simoes M.;
 RT "E-cadherin gene alterations are rare events in thyroid tumors.";
 RL Int. J. Cancer 70:32-38(1997).
 RN [13]
 RP VARIANTS ASP-336 AND ILE-470.
 RX MEDLINE: 98196671.
 RA Guilford P., Hopkins J., Harraway J., McLeod M., McLeod N.,
 RA Harawira P., Taite H., Scoullar R., Miller A.E.,
 RT "E-cadherin germline mutations in familial gastric cancer.";
 RL Nature 392:402-405(1998).
 CC -I- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / SEY6210:
RX MEDLINE: 97060018.
RA Page N., Sheraton J., Brown J.L., Stewart R.S., Bussey H.;
RT "Identification of ASK10 as a multicopy activator of Skn/P-dependent
RL transcription of a HIS3 reporter gene.";
RL Yeast 12:267-272(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Hernandez K., Weber N., Wipfl P., Schmidheini T.;
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PUTATIVE ACTIVATOR OF SKN7.
CC -1- SIMILARITY: TO YEAST YIL105C AND YNL047C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U27209; AAA67368.1; -
DR EMBL: Z72882; CAA97100.1; -
DR SGD: L0002770; ASK10.
FT DOMAIN 22 26 POLY-GLY.
FT DOMAIN 625 628 POLY-SER.
FT DOMAIN 933 938 POLY-ASN.
FT DOMAIN 958 961 POLY-GLN.
FT DOMAIN 972 975 POLY-SER.
FT CONFLICT 57 57 I -> T (IN REF. 1).
FT CONFLICT 346 346 T -> N (IN REF. 1).
FT CONFLICT 464 464 P -> R (IN REF. 1).
FT CONFLICT 467 467 V -> A (IN REF. 1).
FT CONFLICT 603 603 A -> V (IN REF. 1).
FT CONFLICT 906 906 L -> P (IN REF. 1).
SQ SEQUENCE 1146 AA: 126863 MW: 9E9EF22B0EE496EE CRC64:

Query Match 56.3%; Score 54; DB 1; Length 1146;
Best Local Similarity 70.0%; Pred. No. 1.79e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 692 TRKIVSPEPS 701
QY 6 AFKIVSOEPA 15

Search completed: Sat May 13 07:07:11 2000
Job time : 90 secs.

```


THIS PAGE BLANK (uspto)

MUSCLE
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Muscle - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:07:31 2000; Maspar time 140.71 Seconds

Tabular output not generated. 7.391 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table:
PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 25.549; Variance 30.093; scale 0.849

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description	Pred. No.
1	92	95.8	993 11	DESMOGLEIN 3 (FRAGMENT	2.67e-09
2	61	63.5	1421 10	HYPOHETICAL 16.6 KD P	1.77e-01
3	58	60.4	138 2	HYPOHETICAL 16.6 KD P	8.20e-01
4	55	57.3	225 5	ORF M5032 HYPOHETICA	3.62e+00
5	55	57.3	252 14	ORF M5032 HYPOHETICA	3.62e+00
6	55	57.3	327 5	ORF M5032 HYPOHETICA	3.62e+00
7	55	57.3	967 4	ORF M5032 HYPOHETICA	3.62e+00
8	54	56.3	108 10	ORF M5032 HYPOHETICA	3.62e+00
9	54	56.3	238 11	ORF M5032 HYPOHETICA	3.62e+00
10	54	56.3	878 4	ORF M5032 HYPOHETICA	3.62e+00
11	54	56.3	906 11	ORF M5032 HYPOHETICA	3.62e+00
12	54	56.3	2258 14	ORF M5032 HYPOHETICA	3.62e+00
13	54	56.3	2258 14	ORF M5032 HYPOHETICA	3.62e+00
14	53	55.2	464 4	ORF M5032 HYPOHETICA	3.62e+00
15	53	55.2	775 10	ORF M5032 HYPOHETICA	3.62e+00
16	53	55.2	1980 14	ORF M5032 HYPOHETICA	3.62e+00
17	53	55.2	2223 14	ORF M5032 HYPOHETICA	3.62e+00
18	53	55.2	2228 14	ORF M5032 HYPOHETICA	3.62e+00
19	53	55.2	2228 14	ORF M5032 HYPOHETICA	3.62e+00
20	53	55.2	2228 14	ORF M5032 HYPOHETICA	3.62e+00

RESULT	1	PRELIMINARY:	PRT:	993 AA.
AC	035902			
AD	035902			
DT	01-JAN-1998	(TREMBLrel. 05	Created)	
DT	01-JAN-1998	(TREMBLrel. 05	Last sequence update)	
DT	01-NOV-1998	(TREMBLrel. 12	Last annotation update)	
DE	DESMOGLEIN 3 (FRAGMENT)			
GN	DSG3			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
CC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C;			
RA	ISHIKAWA H., LI K., UIRTO J.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).			
DR	EMBL: U86016; AAB65091.1; -			
DR	HSSP: P15116; INCU.			
DR	PROSITE: PS00232; CADHERIN; 2.			
DR	PFAM: PF00028; cadherin; 4.			
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.			
FT	NON_TER 993 993			
SO	SEQUENCE 993 AA; 10788 MW; 881794BD CRC32;			

ALIGNMENTS

RESULT	2	PRELIMINARY:	PRT:	1421 AA.
AC	080907			
DT	01-NOV-1998	(TREMBLrel. 08	Created)	
DT	01-NOV-1998	(TREMBLrel. 08	Last sequence update)	
DT	01-NOV-1998	(TREMBLrel. 08	Last annotation update)	
DE	T19C21.7 PROTEIN.			
GN	T19C21.7			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			

OC euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 RC Arabidopsis.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
 RA SYKES S.M., KAUS S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
 RA SOMERVILLE C.R., VENTER J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence,"
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC004683; AAC28760.1;
 SO SEQUENCE 1421 AA; 154325 MW; 1B6D2FF1 CRC32;
 Query Match 63.5%; Score 61; DB 10; Length 1421;
 Best Local Similarity 40.0%; Pred. No. 1.77e+01;
 Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;
 Db 262 VDSKISYEIITONPG 276
 OY 1 LNSKIARFVISOEPA 15
 RESULT 3
 ID P72670 PRELIMINARY; PRT; 138 AA.
 AC P72670;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 16.6 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.,
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASANIZU E., NAKAMURA Y.,
 RA MIYAIKITA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
 RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
 RA TABATA S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL: D90899; BAA16672.1;
 KW Hypothetical protein.
 SO SEQUENCE 138 AA; 16580 MW; CA436352 CRC32;
 Query Match 60.4%; Score 58; DB 2; Length 138;
 Best Local Similarity 38.5%; Pred. No. 8.20e+01;
 Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
 Db 28 AKLAFEVIPDPT 40
 OY 3 SKIAFKIVISOEPA 15
 RESULT 4
 ID Q20152 PRELIMINARY; PRT; 225 AA.
 AC Q20152;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
 DE F38B7.4 PROTEIN.
 GN F38B7.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LENNARD N.,
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIRKEN L., ROOPRA A., SANDERS D., SHOMKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMER E., STADEN R., SULSTON J.,
 RA THERIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 DR EMBL: Z74033; CA98472.1;
 SO SEQUENCE 225 AA; 26356 MW; 66EA455 CRC32;
 Query Match 57.3%; Score 55; DB 5; Length 225;
 Best Local Similarity 53.3%; Pred. No. 3.62e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Db 205 LASKVSEFVRSQVPS 219
 OY 1 LNSKIARFVISOEPA 15
 RESULT 5
 ID Q9YWS9 PRELIMINARY; PRT; 252 AA.
 AC Q9YWS9;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE ORF MSV032 HYPOTHETICAL PROTEIN.
 GN MSV032.
 OS Melanoplus sanguinipes entomopoxvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
 OC Entomopoxvirus B.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TUCSON;
 RX MEDLINE: 99102612.
 RA AFONSO C.L., TOLMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.,
 RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
 RL J. Virol. 73:533-552(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TUCSON;
 RA AFONSO C.L., TOLMAN E.R., LU Z., OMA E., KUTISH G.F., ROCK D.L.,
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF063866; AAC97615.1;
 SO SEQUENCE 252 AA; 29068 MW; 3A7E442E CRC32;
 Query Match 57.3%; Score 55; DB 14; Length 252;
 Best Local Similarity 58.3%; Pred. No. 3.62e+00;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 202 SKVPFKILOS 213
 OY 3 SKIAFKIVISOEP 14
 RESULT 6
 ID Q93345 PRELIMINARY; PRT; 327 AA.
 AC Q93345;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)


```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 94242050.
RA RIMM D.L., MORROW J.S.;
RT "Molecular cloning of human E-cadherin suggests a novel subdivision of
RL the cadherin superfamily.";
RN Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
RN [2]
RP SEQUENCE OF 333-472 FROM N.A.
RX MEDLINE: 94306394.
RA BECKER K.F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
RA SIEMER J.R., HOFER H.;
RT "E-cadherin gene mutations provide clues to diffuse type gastric
RL carcinoma.";
RN Cancer Res. 54:3845-3852(1994).
RN [3]
RP SEQUENCE OF 261-388 FROM N.A.
RC TISSUE=LIVER;
RA FRIKEN U.H.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: S72492; AAD14108.1; -.
DR EMBL: S72491; AAD14108.1; JOINED.
DR EMBL: L08599; AAA61259.1; -.
DR EMBL: X52379; CA36522.1; -.
DR HSPB: P09603; 1SUH.
DR PROSITE: P500232; CADHERIN; 3.
DR PFAM: PF00028; cadherin; 5.
DR PFAM: PF01049; Cadherin_C_term; 1.
DR PRINTS: PR00205; CADHERIN.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 878 UVOMORULIN.
FT SO SEQUENCE 878 AA; 96741 MW; 8E7F0180 CRC32;

Query Match 56.3%; Score 54; DB 4; Length 878;
Best Local Similarity 53.8%; Pred. No. 5.86e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 293 NAAIATYILSDP 305
Qy 2 NSKIAFKIVSOEP 14

RESULT 11
ID 0921Y3 PRELIMINARY; PRT; 906 AA.
AC 0921Y3;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TESTICULAR N-CADHERIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE: 98187820.
RA CHUNG S.S., MO M.Y., SILVESTRINI B., LEE W.M., CHENG C.Y.;
RT "Rat testicular N-cadherin: its complementary deoxyribonucleic acid
RL cloning and regulation.";
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF097593; AAC83818.1; -.
DR HSPB: P15116; INCI.
DR PROSITE: P500232; CADHERIN; 3.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT SO SEQUENCE 906 AA; 99686 MW; 6C52E299 CRC32;

Query Match 56.3%; Score 54; DB 11; Length 906;
Best Local Similarity 40.0%; Pred. No. 5.86e+00;

```

```

Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

DB 300 LNSGLRYRILSOAPS 314
Qy 1 LNSKIAFKIVSOEPA 15

RESULT 12
ID 089238 PRELIMINARY; PRT; 2258 AA.
AC 089238;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE LARGE PROTEIN.
OS Human parainfluenza virus 3.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GP;
RX MEDLINE: 98436895.
RA OHSAWA K., YAMADA A., TAKEUCHI K., WATANABE Y., MIYATA H., SATO H.;
RT "Genetic characterization of parainfluenza virus 3 derived from guinea
RL pigs.";
DR J. Vet. Med. Sci. 60:919-922(1998).
DR EMBL: AB012132; BAA32575.1; -.
DR PFAM: PF00946; ParamyxV_RNA_pol; 1.
FT SO SEQUENCE 2258 AA; 258800 MW; 8E00950F CRC32;

Query Match 56.3%; Score 54; DB 14; Length 2258;
Best Local Similarity 33.3%; Pred. No. 5.86e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

DB 998 LDRSVLYRIMNOEPG 1012
Qy 1 LNSKIAFKIVSOEPA 15

RESULT 13
ID 081081 PRELIMINARY; PRT; 2258 AA.
AC 081081;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE RNA POLYMERASE.
CN L.
OS Human parainfluenza virus 3.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JS;
RX MEDLINE: 94091054.
RA STOKES A., TIERNEY E.L., SARRIS C.M., MURPHY B.R., HALL S.L.;
RT "The complete nucleotide sequence of two cold-adapted, temperature-
RL sensitive attenuated mutant vaccine viruses (cp12 and cp45) derived
RT from the JS strain of human parainfluenza virus type 3 (PIV3).";
RN Virus Res. 30:43-52(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JS;
RA DUBBIN A.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JS;
RX MEDLINE: 94091054.
RA DUBBIN A.D.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U51116; ABA48690.1; -.
DR PFAM: PF00946; ParamyxV_RNA_pol; 1.
FT SO SEQUENCE 2258 AA; 258927 MW; 4B7A664B CRC32;

Query Match 56.3%; Score 54; DB 14; Length 2258;

```

Best Local Similarity 33.3%: Pred. No. 5.86e+00;
Matches 5: Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 998 LDRSVLYRIMNOEPC 1012

OY 1 LNSKIAFKIVSOEPA 15

RESULT 14

ID 043159

AC 043159

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)

DE KIAA0409 (FRAGMENT).

GN KIAA0409.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RA ISHIIKAWA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,

RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB007869; BAA23705.1; -

FT NON_TER

FT SEQUENCE 464 AA: 51479 MW: 4C0DDCE CRC32:

Query Match

Best Local Similarity 55.2%: Score 53; DB 4; Length 464;

Matches 6: Conservative 4; Mismatches 1; Indels 0; Caps 0;

Db 416 TKLGFKIVSKD 426

OY 3 SKIAFKIVSOE 13

RESULT 15

ID P93205

AC P93205

DT 01-MAY-1997 (TREMBlrel. 03, Created)

DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE SERINE PROTEASE, SBT2.

GN SBT2.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;

Db 734 LSKVTEKTVSR0KA 748

OY 1 LNSKIAFKIVSOEPA 15

Search completed: Sat May 13 07:10:03 2000
Job time : 152 secs.

THIS PAGE BLANK (USPTO)

M P E R E H (TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:16:39 2000: Maspar time 3.02 Seconds
Tabular output not generated. 117,602 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPFMLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 19.380; Variance 58.525; scale 0.331

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	100.0	15	W04844	Self epitope of desmog	6.82e-05
2	108	100.0	15	W64816	Desmoglein 3 206-220.	6.82e-05
3	108	100.0	15	W78815	Desmoglein 3 protein f	6.82e-05
4	108	100.0	614	W07908	Pemphigus vulgaris ant	6.82e-05
5	108	100.0	999	R30742	Human pemphigus vulgar	6.82e-05
6	88	81.5	778	W15489	Pemphigus foliaceus an	1.76e-02
7	61	56.5	583	R07999	Asparagine synthetase	2.03e-01
8	61	56.5	583	R07998	Asparagine synthetase	2.03e-01
9	60	55.6	163	W28654	S. pneumoniae aspartat	2.59e-01
10	59	54.6	616	R68667	Rat pneumocadherin p5.	3.31e-01
11	57	52.8	263	W13010	Segment of desmosomal	5.39e-01
12	57	52.8	560	W13009	Segment of desmosomal	5.39e-01
13	55	50.9	660	R69633	Human interleukin-12 r	8.73e-01
14	55	50.9	662	R69632	Human interleukin-12 r	8.73e-01
15	55	50.9	662	W12772	Human interleukin-12 b	8.73e-01
16	54	50.0	325	W92443	Mouse MC5 protein.	1.11e-02
17	54	50.0	325	W87870	Mouse melanocortin rec	1.11e-02
18	54	50.0	325	W79688	Melanocortin-5 recepto	1.11e-02
19	54	50.0	325	W19705	Melanocortin-5 recepto	1.11e-02
20	54	50.0	325	W33725	Mouse melanocortin-5 (1.11e-02
21	54	50.0	325	R79501	Rat melanocortin recept	1.11e-02
22	54	50.0	325	W41067	Mouse melanocortin-5 r	1.11e-02
23	54	50.0	325	W37833	Mouse melanocortin-5 r	1.11e-02

24	54	50.0	433	1	W00365	Human cyclin B1.	1.11e-02
25	54	50.0	555	1	W85133	A desaturase enzyme en	1.11e-02
26	54	50.0	555	1	W84154	Human desaturase enzym	1.11e-02
27	54	50.0	608	1	W85134	A desaturase enzyme en	1.11e-02
28	54	50.0	608	1	W84155	Human desaturase enzym	1.11e-02
29	54	50.0	746	1	W85135	A desaturase enzyme en	1.11e-02
30	54	50.0	746	1	W84156	Human desaturase enzym	1.11e-02
31	53	49.1	609	1	P30206	Sequence encoded by PF	1.40e-02
32	51	47.2	400	1	R10919	Human GM-CSF receptor.	2.24e-02
33	51	47.2	552	1	R24017	Fusion protein GM-CSF r	2.24e-02
34	51	47.2	1686	1	W70991	Human class II p13 kin	2.24e-02
35	51	47.2	1726	1	W38756	Phosphatidyl inositol	2.24e-02
36	50	46.3	166	1	P70358	Sequence encoded by PI	2.83e-02
37	50	46.3	193	1	P80877	Hook region #8 contig 1	2.83e-02
38	50	46.3	225	1	P70061	Sequence of human grow	2.83e-02
39	50	46.3	297	1	W87867	Bovine melanocortin re	2.83e-02
40	50	46.3	297	1	W92440	Human MC2 protein.	2.83e-02
41	50	46.3	472	1	W40072	Human retinoid recepto	2.83e-02
42	50	46.3	720	1	W19266	Lactobacillus amylovor	2.83e-02
43	50	46.3	878	1	R85487	Human E-cadherin precu	2.83e-02
44	50	46.3	916	1	Y10967	H. pylori ORF hp5e1521	2.83e-02
45	50	46.3	928	1	W21017	H. pylori cytoplasmic	2.83e-02

ALIGNMENTS

RESULT 1
ID W04844; standard; peptide: 15 AA.
AC W04844;
DE 18-FEB-1997 (first entry)
KW Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation, self-epitope; antigen; autoimmune disease.
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphonomomastase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PI (HARD) HARVARD COLLEGE.
PI Strominger JL, Nucheffermly KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease.
PS Claim 1: Page 40; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 206-220)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 TPFMLSRNTGEVRT 15
OY 1 TPFMLSRNTGEVRT 15

RESULT 2
ID W64816 standard; peptide: 15 AA.
AC W64816:
DE 29-SEP-1998 (first entry)
DT Desmoglein-3 206-220.
KW Desmoglein; Dg; gene therapy; pemphigus vulgaris; microparticle;
KW autoantigen; autoimmune disease; MHC.
OS Homo sapiens.
PN US5783567-A.
PM 21-JUL-1998.
PF 22-JAN-1997: 787547.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS;
DR WPI: 98-427077/36.
PT Microparticle encapsulated nucleic acids - for recombinant
PS expression of proteins e.g. in gene therapy.
CC Disclosure; Column 4: 42pp; English.
CC The patent describes a new preparation of microparticles each
CC comprising a polymeric matrix and a nucleic acid. The polymeric
CC matrix consists of one or more synthetic polymers having a solubility
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
CC and at least 90% of the microparticles have a diameter of less than
CC 100 microns. The microparticles are useful for the delivery of nucleic
CC acids to phagocytic cells. In one embodiment the microparticles are
CC less than 20 microns in diameter and the nucleic acid (preferably in
CC closed circular form) includes an expression control sequence
CC operatively linked to a coding sequence, where the expression product
CC of the coding sequence is a polypeptide having a length and a sequence
CC which permits it to bind to an MHC class I or II molecule. The
CC expression product is thus an effective stimulator of an immune
CC response in mammals. The present sequence, an antigenic portion of
CC desmoglein 3, is an example of an MHC class II peptide which can be
CC expressed by the nucleic acid. It is associated with pemphigus
CC vulgaris.
SQ Sequence 15 AA:

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 1TPMFLSRTNGEVRT 15
QY 1 1TPMFLSRTNGEVRT 15

RESULT 3
ID W78815 standard; peptide: 15 AA.
AC W78815:
DE 17-NOV-1998 (first entry)
DT Desmoglein 3 protein fragment 206-220.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
PN WO9831398-A1.
PM 23-JUL-1998.
PF 22-JAN-1998: U01499.
PR 06-JAN-1998: US-003253.
PR 22-JAN-1997: US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR WPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PS matrix and nucleic acid comprising an expression vector for use in
PS gene therapy.
CC Disclosure; Page 8: 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression

CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (1) a
CC fragment of a naturally-occurring mammalian protein; or (11) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 15 AA:

Query Match 100.0%; Score 108; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 1TPMFLSRTNGEVRT 15
QY 1 1TPMFLSRTNGEVRT 15

RESULT 4
ID W07908 standard; protein: 614 AA.
AC W07908:
DE 29-JAN-1997 (first entry)
DT Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; Immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PM 23-JUL-1996.
PF 30-JUN-1995: 165632.
PR 30-JUN-1994: JP-173291.
PA (NISHU) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto-antibody -
PS useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA:

Query Match 100.0%; Score 108; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 6.82e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 205 1TPMFLSRTNGEVRT 219
QY 1 1TPMFLSRTNGEVRT 15

RESULT 5
ID R30742 standard; Protein: 999 AA.
AC R30742:
DE 14-JUN-1993 (first entry)
DT Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.

PN US798918-A.
 PD 15-DEC-1992.
 PF 27-NOV-1991: 798918.
 PR 27-NOV-1991: US-798918.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Amagai M, Klaus-Kovtun V, Stanley JR;
 DR WPI: 93-067436/08.
 N-PSDB: 035992.
 PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
 PT diagnostic and therapeutic uses
 PS Disclosure; Fig 7; 50pp; English.
 CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
 CC and its encoding DNA may be used in the diagnosis and treatment of
 CC pemphigus vulgaris. It is thought that the antigen may be a cell
 CC adhesion molecule.
 SO Sequence 999 AA;

Query Match 100.0%; Score 108; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 6.82e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPMFLSRNTEGVRT 220
 1 TPMFLSRNTEGVRT 15

RESULT 6
 ID W15489 standard; Protein: 778 AA.
 AC W15489;
 DT 17-JUN-1997 (first entry)
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
 KM Pemphigus foliaceus autoantibody; constant region; IgG;
 KW extracellular region; antigen; hinge portion; skin;
 OS Chimeric - Homo sapiens.
 FH Key Location/Qualifiers
 FT domain 1..545
 FT /note= "Pemphigus foliaceus antigen protein"
 PN J09077800-A.
 PD 25-MAR-1997.
 PF 12-SEP-1995: 260899
 PR 12-SEP-1995: JP-260899.
 PA (NISH/) NISHIKAWA T.
 DR WPI: 97-241758/22.
 P-PSDB: T66428.
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
 PT through the hinge region used to treat pemphigus foliaceus
 PS Claim 1: Page 10-12; 17pp; Japanese.
 CC This sequence represents a fused protein recognised by pemphigus
 CC foliaceus patient autoantibody which comprises the constant region
 CC of IgG linked to the extracellular region of pemphigus foliaceus
 CC antigen protein through the hinge portion. Pemphigus foliaceus is
 CC a chronic, generalised, vesicular and scaling skin eruption similar
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
 CC protein is useful to treat pemphigus foliaceus. The antigen is
 CC especially administered through an adsorbent upon which the fusion
 CC protein is immobilised via a carrier. The fusion protein is also
 CC useful for detecting pemphigus foliaceus antibodies which is useful
 CC in immunodiagnosis. The fusion protein has little or no side effects.
 SO Sequence 778 AA;

Query Match 81.5%; Score 88; DB 1; Length 778;
 Best Local Similarity 66.7%; Pred. No. 1.76e-02;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFLIRNTEGVRT 220
 1 TPMFLSRNTEGVRT 15

RESULT 7
 ID R07999 standard; Protein: 583 AA.
 AC R07999;

DT 26-FEB-1991 (first entry)
 DE Asparagine synthetase AS2
 KM asparagine synthetase; transgenic plant; herbicide resistance;
 KW drought tolerance; nitrogen fixation; pea.
 OS Pisum sativum.
 FH Key Location/Qualifiers
 FT binding-site 1..4
 FT /label= glutamine binding site
 FT /note= "conserved in human AS and pea AS1 and 2."
 PN W09013633-A.
 PD 15-NOV-1990.
 PF 02-MAY-1990: U02443.
 PR 03-MAY-1989: US-347302.
 PR 26-APR-1990: US-514816.
 PA (YURO-) ROCKFELLER UNIV.
 PI Coruzzi GM, Tsai FY;
 DR WPI: 90-361471/48.
 N-PSDB: 006599.
 PT Plant asparagine synthetase - includes DNA expression system and
 PT transgenic plants
 PS Disclosure; Fig 2B; 91pp; English.
 CC The DNA sequence encoding this protein was isolated from cDNA
 CC clones selected from a pea nodule cDNA library from the "Sparkle"
 CC variety of P.sativum. Pea AS1 cDNA was used as a probe.
 CC The protein is produced by expression vectors containing the AS2
 CC coding sequence. Recombinant AS can be used to engineer herbicide
 CC resistance, as a dominant selectable marker, to select novel
 CC herbicides or compounds useful for synchronising plant cells in
 CC culture, etc.
 CC A comparison of pea AS and human AS polypeptides reveals an overall
 CC homology of 47% at the amino acid level. C.f. 86% between AS1 and
 CC AS2. There are several regions of high local homology (greater than
 CC 80%) shared between the pea AS and human AS polypeptides.
 CC See also Q06598, Q06622 and Q06623.
 SO Sequence 583 AA;

Query Match 56.5%; Score 61; DB 1; Length 583;
 Best Local Similarity 77.8%; Pred. No. 2.03e+01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMFLMSRK 330
 1 TPMFLSRN 9

RESULT 8
 ID R07998 standard; Protein: 586 AA.
 AC R07998;
 DT 26-FEB-1991 (first entry)
 DE Asparagine synthetase AS1.
 KM asparagine synthetase; transgenic plant; herbicide resistance;
 KW drought tolerance; nitrogen fixation; pea.
 OS Pisum sativum.
 FH Key Location/Qualifiers
 FT region 116..128
 FT /label= 80% homology with human AS1 sequence
 FT region 218..243
 FT /label= 80% homology with human AS1
 FT region 340..348
 FT /label= 80% homology with human AS1
 FT region 352..360
 FT /label= 80% homology with human AS1
 FT region 392..401
 FT /label= 80% homology with human AS1
 FT region 486..500
 FT /label= 80% homology with human AS1
 FT binding-site 1..4
 FT /label= putative glutamine binding site
 PN W09013633-A.
 PD 15-NOV-1990.
 PF 02-MAY-1990: U02443.
 PR 03-MAY-1989: US-347302.
 PR 26-APR-1990: US-514816.

PA (UVR0-) ROCKFELLER UNIV.
 PI CORUZ21 GM, Tsal FY;
 DR WPI: 90-361471/48.
 DR N-PSDB: 006599.
 PT Plant asparagine synthetase - includes DNA expression system and
 PS transgenic plants
 PS Disclosure: fig 2A: 91pp: English.
 CC The DNA sequence encoding this protein was isolated from cDNA
 CC clones selected from a pea node cDNA library from the "Sparkle"
 CC variety of *P. sativum*. Human AS cDNA was used as a probe.
 CC The protein is produced by expression vectors containing the AS1
 CC coding sequence. Recombinant AS can be used to engineer herbicide
 CC resistance, as a dominant selectable marker, to select novel
 CC herbicides or compounds useful for synchronising plant cells in
 CC culture, etc.
 CC See also 006599, 006622 and 006623.
 SO Sequence 586 AA;

Query Match 56.5%; Score 61; DB 1; Length 586;
 Best Local Similarity 77.8%; Pred. No. 2.03e+01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 323 TPFLLMSRK 331
 OY 1 TPFLLSRN 9
 ID W38654 standard; Protein: 163 AA.
 AC W38654;
 DT 09-NOV-1998 (first entry)
 DE S. pneumoniae aspartate ammonia lyase.
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KM immunological response; inoculation; antibody production; inhibitor;
 KM T cell immune response; antimicrobial compound; bacterial adhesion;
 KM extracellular matrix protein; protein-mediated cell invasion; wound;
 KM pathogenesis.
 OS Streptococcus pneumoniae.
 PN MO9743303-A1.
 PD 20-NOV-1997.
 PF 14-MAY-1997: U07950.
 PR 14-MAY-1996: US-017670.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI (SMK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Stodola RK;
 DR WPI: 98-008793/01.
 DR N-PSDB: T98699.
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 PS Claim 12: Page 407: 483pp: English.
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with an Escherichia coli protein, is a aspartate ammonia
 CC lyase, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 010093 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 SO Sequence 163 AA;

Query Match 55.6%; Score 60; DB 1; Length 163;
 Best Local Similarity 60.0%; Pred. No. 2.59e+01;
 Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

DB 115 AMFLKRHRICEVQT 129
 OY 2 PMFLLS-RNTGSEVRT 15
 ID R86867 standard; Protein: 616 AA.
 AC R86867;
 DT 27-AUG-1996 (first entry)
 DE Rat protocadherin pc5.
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KM catenin; therapy.
 OS Rattus rattus.
 PN MO9600289-A1.
 PD 04-JAN-1996.
 PF 26-JUN-1995: U08071.
 PR 27-JUN-1994; US-268161.
 PA (DOHE-) DOHERTY EYE INST.
 PI Suzuki S;
 DR WPI: 96-068873/07.
 DR N-PSDB: T03574.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PS Claim 17; Page 130-132: 146pp: English.
 CC R86865-R86867 represent the sequences for three protocadherins. This
 CC sequence represents the rat protocadherin pc5. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherin adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to
 CC function via a different method from those with a cytoplasmic domain.
 CC These sequences were isolated using primers 1 and 2 (see T03575 and
 CC T03576) The proteins may have regulatory functions in the cell, as well
 CC as the cell-cell adhesive properties. Antibodies produced against these
 CC sequences are useful for modulating the binding activity of these
 CC protocadherins, and can be used therapeutically.
 SO Sequence 616 AA;

Query Match 54.6%; Score 59; DB 1; Length 616;
 Best Local Similarity 61.5%; Pred. No. 3.31e+01;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 85 TPLFOLNTEGEI 97
 OY 1 TPFLLSRNTEV 13
 ID W13010 standard; Protein: 263 AA.
 AC W13010;
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2.
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
 KW microstereotaxis; separation; enrichment; targeted delivery;
 KW metastatic.
 OS Homo sapiens.
 PN DE19531033-A1.
 PD 27-FEB-1997.
 PF 23-AUG-1995: 031033.
 PR 23-AUG-1995: DE-031033.
 PA (PROG-) PROGEM BIOTECHNIK GMBH.
 PI Franke WW, Schaefer S;
 DR WPI: 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases
 PS Claim 9: Page 5: 8pp: German.
 CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinomas.
 CC Sequence 263 AA:

Query Match 52.8%; Score 57; DB 1; Length 263;
 Best Local Similarity 50.0%; Pred. No. 5.39e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 58 PFVLYNKDGET 69
 Oy 2 PMFLSRNTGEV 13

RESULT 12
 ID W13009 standard; protein: 560 AA.

AC W13009:
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2.
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial; carcinoma; desmosome; antibody; epitope; diagnosis; detection; micrometastasis; separation; enrichment; targeted delivery;
 KM metastatic.
 OS Homo sapiens.
 PN DE19531033-A1.
 PD 27-FEB-1997.
 PE 23-AUG-1995; 031033.
 PR 23-AUG-1995; DE-031033.
 PA (PROG-) PROGEN BIOTECHNIK GMBH.
 PI Franke WW, Schaefer S;
 DR WPI: 97-146518/14.
 PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases
 PS Claim 7: Page 5: 8pp: German.
 CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinomas.
 CC Sequence 560 AA:

Query Match 52.8%; Score 57; DB 1; Length 560;
 Best Local Similarity 50.0%; Pred. No. 5.39e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 160 PFVLYNKDGET 171
 Oy 2 PMFLSRNTGEV 13

RESULT 13
 ID R69633 standard; protein: 660 AA.

AC R69633:
 DT 05-SEP-1995 (first entry)
 DE Human interleukin-12 receptor.
 KW Interleukin-12 receptor; IL-12; immune suppression; immunosuppressive; graft-versus-host reaction; allograft rejection;
 KM inflammation; autoimmune disease.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT Peptide 1..23
 FT /label- Sig_peptide
 FT region 24..540
 FT /note- "extracellular region"
 FT region 52..64
 FT /note- "sequence motif of cytokine receptor superfamily Cys52..Cys62sw"
 FT modified_site 121..123
 FT /label- N-linked_glycosylation
 FT region 222..226
 FT /note- "cytokine receptor superfamily motif (W225KWS)"
 FT modified_site 329..331
 FT /label- N-linked_glycosylation
 FT modified_site 346..348
 FT /label- N-linked_glycosylation
 FT modified_site 352..354
 FT /label- N-linked_glycosylation
 FT modified_site 442..444
 FT /label- N-linked_glycosylation
 FT modified_site 456..458
 FT /label- N-linked_glycosylation
 FT region 541..570
 FT /label- Transmembrane_region
 FT region 571..562
 FT /label- Cytoplasmic_tail_region
 FT region 577..584
 FT /label- Cytoplasmic_tail
 FT /note- "conserved area of cytoplasmic tail"
 FT region 618..629
 FT /label- Cytoplasmic_tail
 FT /note- "conserved area of cytoplasmic tail"

FT EP-638644-A.
 FT PD 15-FEB-1995.
 PE 08-JUL-1994; 110657.
 PR 19-JUL-1993; US-094649.
 PR 31-MAY-1994; US-094713.
 PA (HOFF-) HOFFMANN LA ROCHE & CO AG F.
 PI Chizzonite RA, Chua AO, Gubler UA, Trullt TP;
 DR WPI: 95-076349/11.
 PT DNA encoding a low affinity interleukin-12 receptor - used to bind or scavenge IL-12 to cause immune suppression, e.g. to suppress graft-vs-host reaction, allograft rejection or inflammation, and to treat autoimmune conditions
 PS Claim 4: Page 31-33; 61pp: English.
 CC A cDNA library of PHA-activated peripheral blood mononuclear cells in vector per-BOS was screened for interleukin-12 (IL-12) receptor cDNAs by panning. One isolated cDNA (083844) encoded a 662-amino acid low affinity IL-12 receptor (R69632). Another cDNA contained an additional 202 bp of 3'UTR; the deduced 660-amino acid sequence (R69633) showed an altered C-terminal sequence compared to R69632.
 CC Recombinant IL-12 receptor was expressed in COS cells, and can be used for therapeutic or diagnostic purposes.
 CC Sequence 660 AA:

Query Match 50.9%; Score 55; DB 1; Length 660;
 Best Local Similarity 53.8%; Pred. No. 8.73e+01;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLLSRGACACT 26
 Oy 3 MFLSRNTGEV 15

RESULT 14
ID R69632 standard: Protein: 662 AA.
AC R69632;
DT 05-SEP-1995 (first entry)
DE Human Interleukin-12 receptor.
KW Interleukin-12 receptor; IL-12; Immune suppression;
KW Immunosuppressive; graft-versus-host reaction; allograft rejection;
KW Inflammation; autoimmune disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..23
FT /label= Sig-peptide
FT region 24..540
FT /note= "extracellular region"
FT region 52..64
FT /note= "sequence motif of cytokine receptor
FT superfamily Cys52..Cys62SW"
FT modified_site 121..123
FT /label= N-linked-glycosylation
FT region 222..226
FT /note= "cytokine receptor superfamily motif
FT (W22SKWS)"
FT modified_site 329..331
FT /label= N-linked-glycosylation
FT modified_site 346..348
FT /label= N-linked-glycosylation
FT modified_site 352..354
FT /label= N-linked-glycosylation
FT modified_site 442..444
FT /label= N-linked-glycosylation
FT modified_site 456..458
FT /label= N-linked-glycosylation
FT region 541..570
FT /label= Transmembrane-region
FT region 571..662
FT /label= Cytoplasmic-tail-region
FT region 577..584
FT /label= Cytoplasmic-tail
FT /note= "conserved area of cytoplasmic tail"
FT region 618..629
FT /label= Cytoplasmic-tail
FT /note= "conserved area of cytoplasmic tail"
PN EP-638644-A.
PN 15-FEB-1995.
PF 08-JUL-1994: 110657.
PR 19-JUL-1993: US-094649.
PR 19-JUL-1993: US-094713.
PR 31-MAY-1994: US-248532.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Chizzonite RA, Chua AO, Gubler UA, Trutt TP;
DR WPI: 95-076349/11.
DR N-PSDB: 083844.
PT DNA encoding a low affinity Interleukin-12 receptor - used to
PT bind or scavenge IL-12 to cause immune suppression, e.g. to
PT suppress graft-vs-host reaction, allograft rejection or
PT inflammation, and to treat autoimmune conditions
PS Claim 4: Page 24-27: 61pp: English.
CC A cDNA library of PHA-activated peripheral blood mononuclear cells
CC in vector pEF-BOS was screened for Interleukin-12 (IL-12) receptor
CC cDNAs by panning. An isolated cDNA was sequenced (083844); it
CC encoded a 662-amino acid low affinity IL-12 receptor (R69632).
CC Recombinant IL-12 receptor was expressed in COS cells, and can be
CC used for therapeutic or diagnostic purposes.
SQ Sequence 662 AA;

Query Match 50.9%; Score 55; DB 1; Length 662;
Best Local Similarity 53.8%; Pred. No. 8.73e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLSRGACRT 26
OY 3 MFLSRNTGEVRT 15

RESULT 15
ID W12772 standard: Protein: 662 AA.
AC W12772;
DT 12-MAY-1997 (first entry)
DE Human Interleukin-12 beta-1 receptor.
KW Interleukin-12 beta-1 receptor; IL-12; autoimmune disease;
KW rheumatoid arthritis; inflammatory bowel disease;
KW multiple sclerosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..20
FT /label= Sig-peptide
FT /note= "signal peptide cleavage site alternatively
FT follows Ala-23 in Cys-24"
FT domain 24..540
FT /label= Extracellular-domain
FT domain 540..570
FT /label= Transmembrane-domain
FT domain 571..662
FT /label= Cytoplasmic-tail
FT region 52..64
FT /label= Cytokine-receptor-motif
FT /note= "cytokine receptor superfamily motif
FT (Cys52..Cys62SW)"
FT region 222..226
FT /label= Cytokine-receptor-motif
FT /note= "cytokine receptor superfamily motif
FT (W22SKWS)"
FT modified_site 121..123
FT /label= N-linked-glycosylation-site
FT modified_site 329..331
FT /label= N-linked-glycosylation-site
FT modified_site 346..348
FT /label= N-linked-glycosylation-site
FT modified_site 352..354
FT /label= N-linked-glycosylation-site
FT modified_site 442..444
FT /label= N-linked-glycosylation-site
FT modified_site 456..458
FT /label= N-linked-glycosylation-site
PN EP-759466-A2.
PN 26-FEB-1997.
PF 23-JUL-1996: 111807.
PR 30-AUG-1995: US-001701.
PR 30-MAY-1996: US-018674.
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
PI Gubler UA, Presky DH;
DR WPI: 97-147515/14.
DR N-PSDB: T59732.
PT New Interleukin-12 beta-2 receptor and high binding affinity
PT complexes - have a high affinity for Interleukin-12, and are used to
PT treat autoimmune diseases
PS Claim 8: Page 37-41: 53pp: English.
CC Human Interleukin-12 (IL-12) receptor beta-1 receptor (W12772)
CC has a low binding affinity for IL-12, but when complexed with an
CC IL-12 beta-2 receptor (see also W12771), forms a complex with a
CC high binding affinity for IL-12. Its amino acid sequence was
CC deduced from a cDNA clone (T59732) obtd. from human lymphoblasts.
CC IL-12 receptor beta-1 can be expressed on the surface of
CC transformed host cells as a complex with co-expressed IL-12 receptor
CC beta-2, and used in therapeutic compns.. pref. with at least 1
CC cytokine antagonist, to treat autoimmune dysfunctions such as
CC rheumatoid arthritis, inflammatory bowel disease and multiple
CC sclerosis. The receptor protein or complex can also be used to
CC detect (ant)agonists of IL-12 activity.
SQ Sequence 662 AA;

Query Match 50.9%; Score 55; DB 1; Length 662;
Best Local Similarity 53.8%; Pred. No. 8.73e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 14 LFLSRGACRT 26

Mon May 15 08:08:39 2000

US-08-991-628-4a.rag

Page 7

QY :||||| :|||
3 MFLSRNTGEVRT 15

Search completed: Sat May 13 07:16:47 2000
Job time : 8 secs.

[W][O][R][D] (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:16:14 2000; MasPar time 4.29 Seconds
Tabular output not generated. 165.096 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSRMTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 27.219; Variance 35.537; scale 0.766

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	100.0	999	1 IJHUG3	desmoglein 3 precursor	1.65e+11
2	88	81.5	1049	1 IJHUG1	desmoglein 1 precursor	1.37e+06
3	81	73.0	1043	1 IJHUG1	desmoglein 1 precursor	5.83e-05
4	65	60.2	572	2 S52694	asparagine synthase (1.74e-01
5	65	60.2	572	2 S55982	asparagine synthase (1.74e-01
6	63	58.3	295	2 S32892	hypothetical protein	4.41e-01
7	63	58.3	295	2 S56364	inner membrane copper	4.41e-01
8	63	58.3	923	2 S50482	hypothetical protein	4.41e-01
9	61	56.5	581	2 T08846	asparagine synthase (1.10e+00
10	61	56.5	583	1 AJPMN2	asparagine synthase (1.10e+00
11	61	56.5	584	2 T12989	asparagine synthase (1.10e+00
12	61	56.5	586	2 S69182	asparagine synthase (1.10e+00
13	61	56.5	586	2 S69183	asparagine synthase (1.10e+00
14	61	56.5	586	1 AJPMN1	asparagine synthase (1.10e+00
15	61	56.5	586	2 T02978	asparagine synthase (1.10e+00
16	61	56.5	586	2 S52387	asparagine synthase (1.10e+00
17	61	56.5	591	2 T03602	probable asparagine s	1.10e+00
18	59	54.6	148	1 E71058	hypothetical protein	2.68e+00
19	59	54.6	163	2 B75085	iron-dependent repres	2.68e+00
20	59	54.6	590	2 S25165	asparagine synthase (2.68e+00
21	59	54.6	852	2 B72685	hypothetical protein	2.68e+00
22	59	54.6	1102	2 JH0717	guanylate cyclase (EC	2.68e+00
23	58	53.7	593	2 T10301	ribonucleotide reduct	4.16e+00

24	57	52.8	171	1 I36845	dual specificity phos	6.42e+00
25	57	52.8	483	2 T06459	62k sucrose-binding p	6.42e+00
26	57	52.8	554	1 AJECN	asparagine synthase (6.42e+00
27	57	52.8	579	2 A64100	inner membrane copper	6.42e+00
28	57	52.8	1117	2 S36873	desmoglein 2 - human	6.42e+00
29	56	51.9	276	2 B72582	probable nodulation A	9.84e+00
30	56	51.9	496	2 D71089	hypothetical protein	9.84e+00
31	56	51.9	579	2 T00071	asparagine synthase (9.84e+00
32	55	50.9	124	2 G70785	hypothetical protein	1.50e+01
33	55	50.9	186	2 S04671	H+-transporting ATP s	1.50e+01
34	55	50.9	366	2 F70952	probable corA protein	1.50e+01
35	55	50.9	367	2 T02000	hypothetical protein	1.50e+01
36	55	50.9	375	1 YKMT	citrate (31)-synthase	1.50e+01
37	55	50.9	662	2 I37892	IL12 receptor compone	1.50e+01
38	55	50.9	740	2 S17925	polynucleotide adenyl	1.50e+01
39	54	50.0	152	2 S36550	E6 protein - human pa	2.28e+01
40	54	50.0	163	2 A71536	hypothetical protein	2.28e+01
41	54	50.0	244	1 S16307	transcription activat	2.28e+01
42	54	50.0	325	2 JC2193	melanocortin receptor	2.28e+01
43	54	50.0	433	2 A32892	cyclin B1 - human	2.28e+01
44	54	50.0	582	2 A70755	hypothetical protein	2.28e+01
45	54	50.0	832	2 H72278	alpha-mannosidase-rel	2.28e+01

ALIGNMENTS

RESULT	ENTRY	1	ALIGNMENTS
ENTRY	IJHUG3	#type complete	
TITLE	desmoglein 3 precursor - human		
ALTERNATE_NAMES	pemphigus vulgaris antigen		
ORGANISM	Homo sapiens	#common-name man	
DATE	30-Jun-1993	#sequence-revision 30-Jun-1993	#text-change

22-Jun-1993

ACCESSIONS	A41088
REFERENCE	A41088
#authors	Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal	Cell (1991) 67:865-877
#title	Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.
#cross-references	MUID:92069753
#accession	A41088
#molecule-type	mRNA
#residues	1-999
#label	AMA
#cross-references	GB:M76482; NID:q190751; PIDN:AAA60230.1; PID:q190752

GENETICS

#gene	GDB:DSG3
#cross-references	GDB:134030; OMIM:169615
#map-position	18q12.1-18q12.2
CLASSIFICATION	#superfamily cadherin; cadherin repeat homology
KEYWORDS	calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE

1-23	#domain signal sequence	#status predicted	#label SIG
24-49	#domain propeptide	#status predicted	#label PRO
50-999	#product desmoglein homolog	#status predicted	#label
	MAT		
50-615	#domain extracellular	#status predicted	#label EXT
52-157	#domain cadherin repeat	#status predicted	#label CR1
160-667	#domain cadherin repeat	#status predicted	#label CR2
270-383	#domain cadherin repeat	#status predicted	#label CR3
390-495	#domain cadherin repeat	#status predicted	#label CR4
496-598	#domain cadherin repeat	#status predicted	#label CR5
616-839	#domain transmembrane	#status predicted	#label TM
640-999	#domain intracellular	#status predicted	#label INT
910-938	#domain desmoglein repeat	#label DSG1	
937-966	#domain desmoglein repeat	#label DSG2	
110,180,545	#binding-site carbohydrate (asn)	(covalent)	#status predicted

SUMMARY #length 999 #molecular-weight 107502 #checksum 8311

Query Match 100.0%; Score 108; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.65e+11;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPFLLSRNNGEVRT 220
 Oy 1 TPFLLSRNNGEVRT 15

RESULT 2

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein 1
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S16906; A39706; A61254; A61279; S16158

REFERENCE
 #authors Buxton, R.S.
 #submission submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 ##label BUX

REFERENCE
 #cross-references EMBL:X56654; NID:g30505; PIDN:CAA39976.1; PID:g30506
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 ##label WHE

REFERENCE
 #cross-references GB:X56654
 #accession A61254
 #molecule_type mRNA
 #residues 26-1049 ##label NIL

REFERENCE
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily.

#cross-references MUID:92175187
 #accession A61279
 #molecule_type mRNA
 #residues 26-1049 ##label NIL

REFERENCE
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily.

#cross-references MUID:92175187
 #accession A61279
 #molecule_type mRNA
 #residues 26-1049 ##label NIL

GENETICS
 #gene GDB:DSG1
 #cross-references GDB:126563; OMIM:125670
 #map_position 18q12.1-18q12.2
 CLASSIFICATION #superfamily cadherin; cadherin repeat homology
 KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
 1-23 #domain signal sequence #status predicted #label SIG
 24-49 #domain propeptide #status predicted #label PRO
 50-1049 #product desmoglein #status predicted #label MAT
 50-548 #domain extracellular #status predicted #label EXT
 52-157 #domain cadherin repeat homology #label CR1
 160-269 #domain cadherin repeat homology #label CR2
 272-385 #domain cadherin repeat homology #label CR3
 392-493 #domain cadherin repeat homology #label CR4
 509-530 #region serine/threonine-rich
 549-569 #domain transmembrane #status predicted #label TMN

572-1049
 840-869 #domain intracellular #status predicted #label INT
 870-899 #domain desmoglein repeat #label DG1
 900-927 #domain desmoglein repeat #label DG2
 928-956 #domain desmoglein repeat #label DG3
 969-1019 #domain desmoglein repeat #label DG4
 110,180 #region glycoprotein repeat #label DG4
 #binding-site carbohydrate (Asn) (covalent) #status predicted

SUMMARY
 #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 81.5% Score 88; DB 1; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 1,37e-06;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFIRNNGEVRT 220
 Oy 1 TPFLLSRNNGEVRT 15

RESULT 3

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS S14603; A38872; A37785; S38721; A48173; S24412

REFERENCE
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submission submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 ##label KOC

REFERENCE
 #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87;968-1043 ##label KOZ

REFERENCE
 #cross-references GB:S64268; GB:S64270
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Rasz, L.; Menabe, M.; Cowin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123, 'V', 125-493 ##label GOO

REFERENCE
 #cross-references GB:M8165; NID:g162966; PIDN:AAA62709.1; PID:g552318
 #accession S38721
 #authors Zimbelmann, R.
 #submission submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 ##label ZIM

REFERENCE
 #cross-references EMBL:X57784; NID:g436061; PIDN:CAA40930.1; PID:g436062
 #authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal

glycoprotein, as a member of the cadherin family of cell adhesion molecules.
#accession MUID:91168965
#molecule_type mRNA
#residues 44-1001, 'AOPSPAR' ##label KO3
##cross-references GB:X57784
##note this sequence has been revised in references A38872 and S38721

GENETICS
#gene DSG1
CLASSIFICATION
#superfamily cadherin; cadherin repeat homology
#calcium binding; cell adhesion; duplication; glycoprotein;
#transmembrane protein

FEATURE
1-23
24-49 #domain signal sequence #status predicted #label SIG
50-1043 #domain propeptide #status predicted #label PRO
50-548 #product desmoglein #status predicted #label MAT
52-157 #domain extracellular #status predicted #label EXT
160-269 #domain cadherin repeat homology #label CR1
392-491 #domain cadherin repeat homology #label CR2
372-385 #domain cadherin repeat homology #label CR3
549-574 #domain transmembrane #status predicted #label TM
575-1043 #domain intracellular #status predicted #label INT
846-875 #domain desmoglein repeat #label DG1
876-905 #domain desmoglein repeat #label DG2
906-933 #domain desmoglein repeat #label DG3
934-962 #domain desmoglein repeat #label DG4
963-1012 #region glycine/serine-rich
110 #binding_site carbohydrate (asn) (covalent) #status experimental
180,496 #binding_site carbohydrate (asn) (covalent) #status predicted

SUMMARY
#length 1043 #molecular-weight 112242 #checksum 6897

Query Match 75.0% Score 81; DB 1; Length 1043;
Best Local Similarity 60.0%; Pred. No. 5,83e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 206 SPMFLINRYTGEIRT 220
1 TPMFLSRNTEGVRT 15

RESULT 4
ENTRY S52694 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN1
ALTERNATE_NAMES - yeast (Saccharomyces cerevisiae)
ORGANISM protein P9659.3; protein YPR145W
#formal_name Saccharomyces cerevisiae
DATE 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 16-Jul-1999
S52694; S69033

ACCESSIONS
REFERENCE S52694
#authors Dang, V.D.; Bolotin-Fukuhara, M.; Dalgman-Fornier, B.
#submission submitted to the EMBL Data Library, March 1995
#description Multiple regulatory systems control expression of the Saccharomyces cerevisiae ASN1 and ASN2 genes at the transcription level.
#accession S52694
#molecule_type DNA
#residues 1-572 ##label DAN
##cross-references EMBL:Z48675; NID:g1163061; PIDN:CAA8594.1; PID:g747902

REFERENCE S69033
#authors Fulton, L.
#submission submitted to the EMBL Data Library, November 1995
#description The sequence of S. cerevisiae cosmid 9659.
#accession S69033
#molecule_type DNA
#residues 1-572 ##label FUL
##cross-references EMBL:U40829; NID:g1066476; PIDN:AAB68284.1;

GENETICS
#gene SGD:ASN1
##cross-references SGD:S0006349; MIPS:YPR145W
#map_position 16R
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE
2
SUMMARY
#active-site Cys (amide transfer) #status predicted
#length 572 #molecular-weight 64470 #checksum 8736

Query Match 60.2% Score 65; DB 2; Length 572;
Best Local Similarity 88.9%; Pred. No. 1.74e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 347 TPMFLSRK 355
1 TPMFLSRN 9

RESULT 5
ENTRY S55982 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN2
ALTERNATE_NAMES - yeast (Saccharomyces cerevisiae)
ORGANISM protein G6358; protein YGR124W
#formal_name Saccharomyces cerevisiae
DATE 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 16-Jul-1999
S55982; S64433

ACCESSIONS
REFERENCE S55976
#authors van Dyck, L.; Goffeau, A.
#submission submitted to the EMBL Data Library, December 1994
#description Genes for an asn synthase, a GUG-motif nucleoporin and a putative homeobox-domain protein are identified on a 18.3 kb segment of the yeast chromosome VII also carrying MEPI, PPT1, tree new ORFs, remnants of Ty and three tRNA genes.
#accession S55982
#molecule_type DNA
#residues 1-572 ##label VAN
##cross-references EMBL:X83099; NID:9642340; PIDN:CAA58159.1; PID:9642347

REFERENCE S64428
#authors Van Dyck, L.; Skala, J.; de Wergifosse, P.; Purnelle, B.; Talle, E.; Nawrocki, A.; Del Bino, S.; Goffeau, A.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64433
#molecule_type DNA
#residues 1-572 ##label VAM
##cross-references EMBL:Z72909; NID:g1323202; PIDN:CAA97135.1; PID:e243463; PID:g1323203; MIPS:YGR124W
##experimental_source strain S288C

GENETICS
#gene SGD:ASN2
##cross-references SGD:S0003356; MIPS:YGR124W
#map_position 7R
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE
2
SUMMARY
#active-site Cys (amide transfer) #status predicted
#length 572 #molecular-weight 64593 #checksum 9651

Query Match 60.2% Score 65; DB 2; Length 572;
Best Local Similarity 88.9%; Pred. No. 1.74e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 346 TPMFLSRK 354
1 TPMFLSRN 9

RESULT 6
ENTRY S33892 #type complete
TITLE hypothetical protein 6 (peti 5' region) - Salmonella

```

ORGANISM      typhimurium
               #formal_name salmonella typhimurium
DATE          19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
               20-Sep-1999

ACCESSIONS
REFERENCE     #authors
              Friedlrich, M.J.; Kinsey, N.E.; Villa, J.; Kadner, R.J.
#journal      Mol. Microbiol. (1993) 8:543-558
#title        Nucleotide sequence of a 13.9kb segment of the 90kb virulence
              plasmid of Salmonella typhimurium: the presence of fimbrial
              biosynthetic genes.

#cross-references EMBL:93316852
#accession      S32892
#status         preliminary
#molecule_type DNA
#residues       1-295 #label FRI
#cross-references EMBL:L08613
CLASSIFICATION #superfamily Salmonella typhimurium hypothetical protein 6
               (pepf 5' region)
SUMMARY        #length 295 #molecular-weight 31162 #checksum 8332

Query Match    58.3% Score 63; DB 2; Length 295;
Best Local Similarity 80.0%; Pred. NO. 4,41e-01;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 132 TLMFSLR-TGEVRT 145
      | | | | | | | | |
Oy 1 TPFLLSRNTGEVRT 15

RESULT 7
ENTRY   S56364 #type complete
TITLE   inner membrane copper tolerance protein cycZ - Escherichia
        coli
ALTERNATE_NAMES
ORGANISM  thio:disulfide interchange protein dsbd
        #formal_name Escherichia coli
DATE      28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
        21-Nov-1997

ACCESSIONS
REFERENCE  S56364; I41028; I41037; S42064; F65223; S57220; S47295
          S56314
#authors  Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;
          Blatter, F.R.
#journal  Nucleic Acids Res. (1995) 23:2105-2119
#title    Analysis of the Escherichia coli genome VI: DNA sequence of
          the region from 92.8 through 100 minutes.
#cross-references EMBL:9534362
#accession  S56364
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-565 #label BUR
#cross-references EMBL:U14003; NID:g1263172; PID:g536980
#note       the nucleotide sequence was submitted to the EMBL Data
          Library, August 1994

REFERENCE     I41027
#authors      Fong, S.T.; Camakaris, J.; Lee, B.T.
#journal      Mol. Microbiol. (1995) 15:1139-1150
#title        Molecular genetics of a chromosomal locus involved in copper
              tolerance in Escherichia coli K-12.
#cross-references EMBL:95349397
#accession    I41028
#status       translated from GB/EMBL/DBJ
#molecule_type DNA
#residues     'M', 78-565 #label RES
#cross-references EMBL:Z36905; NID:g535290; PID:g581055
#note         in this report, the codon GTG for Val-77 was interpreted
              as a start codon

REFERENCE     I41036
#authors      Crooke, H.; Cole, J.
#journal      Mol. Microbiol. (1995) 15:1139-1150
#title        The biosynthesis of c-type cytochromes in Escherichia coli
              requires a membrane-bound protein, DtpZ, with a protein
              disulphide isomerase-like domain.
#cross-references EMBL:95349398

```

```

#accession      141037
##status        translated from GB/EMBL/DBD
#molecule-type DNA
##residues      'M',78-565 ##label RE2
##cross-references EMBL:X77707; NID:9871027; PID:9871029
##note          In this report, the codon GNG for Val-77 was interpreted
                  as a start codon
REFERENCE
#authors        SA2063
#submissions    CTOOKE, H.R.; COLE, J.A.
#description    Submitted to the EMBL Data Library, February 1994
                  The biogenesis of G-type cytochromes in Escherichia coli
                  requires an integral membrane protein with a protein
                  disulphide isomerase like domain.
#accession      SA2064
#molecule-type DNA
##residues      1-328, 'V',330-565 ##label CRO
##cross-references EMBL:X77707
REFERENCE
#authors        Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                  Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                  Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                  Klinkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                  Y.
#journal        Science (1997) 277:1453-1462
#title          The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession      F65223
##status        nucleic acid sequence not shown; translation not shown
#molecule-type DNA
##residues      1-565 ##label BLAT
##cross-references GB:AE000486; GB:U00096; NID:91790574; PID:91790578;
                  UMG: B4136
##experimental_source strain K-12, substrain MG1655
REFERENCE
#authors        Missiakas, D.; Schaefer, F.; Raine, S.
#journal        EMBO J. (1995) 14:3415-3424
#title          Identification and characterization of a new disulfide
                  isomerase-like protein (DsbD) in Escherichia coli.
#cross-references MUID:95354659
#accession      S57220
##status        Preliminary
#molecule-type DNA
##residues      462-540 ##label MIS
GENETICS
#gene           dsbD; cycZ; Cuta2; dlpZ
#keywords        inner membrane; redox-active disulfide; transmembrane protein
#summary        length 565 #molecular-weight 61795 #checksum 2685
Query Match      58.3%; Score 63; DB 2; Length 565;
Best Local Similarity 53.8%; Pred. No. 4; Ale-01;
Matches          7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db              373 PYELLERVIGDWM 385
Oy              2 PYELLSRMTGEVR 14
RESULT          8
ENTRY TITLE     S50482 #type complete
                  hypochlorite protein YER024W - yeast (Saccharomyces
                  cerevisiae)
ORGANISM        #formal_name Saccharomyces cerevisiae
DATE            26-May-1993 #sequence-revision 24-Feb-1995 #text-change
                21-Nov-1997
ACCESSIONS      S50482
REFERENCE        S50482
#authors        Dietrich, F.S.
#submissions    Submitted to the EMBL Data Library, December 1994
#description    The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867,
                  and lambda clone 5698.
#accession      S50482
#molecule-type DNA
##residues      1-923 ##label DIE

```

##cross-references EMBL:U18778; NID:g603592; PID:g603616; MIPS:YER024w
GENETICS
#map-position 5R
SUMMARY #length 923 #molecular-weight 103333 #checksum 767

Query Match 58.3%; Score 63; DB 2; Length 923;
Best Local Similarity 53.8%; Pred. NO. 4.41e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 722 MFMLSVLGEIRS 734
|||||:|||||:
QY 3 MFLLSRNTGEVRT 15

RESULT 9
ENTRY T08846 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
soybean
ALTERNATE_NAMES asparagine synthetase 2
ORGANISM #formal_name Glycine max #common_name soybean
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999

ACCESSIONS
REFERENCE T08846
#authors Hughes, C.A.; Beard, H.; Matthews, B.F.
#journal Plant Mol. Biol. (1997) 33:301-311
#title Molecular cloning and expression of two cDNAs encoding
asparagine synthetase in soybean.
#accession T08846
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-581 #label HUG
##cross-references EMBL:U77678; NID:g1778370
##experimental_source strain Century; leaf

GENETICS
#gene AS2
FUNCTION #description catalyzes ATP-dependent transfer of the amide group of
glutamine to aspartate generating asparagine and glutamate
#pathway asparagine biosynthesis
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 581 #molecular-weight 65609 #checksum 7265

Query Match 56.5%; Score 61; DB 2; Length 581;
Best Local Similarity 77.8%; Pred. NO. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMLMSRK 330
|||||:|||||:
QY 1 TPMLLSRN 9

RESULT 10
ENTRY AUPM2 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
garden pea
ALTERNATE_NAMES asparagine synthetase (glutamine-hydrolyzing)
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
18-Jun-1999

ACCESSIONS
REFERENCE S11443
#authors Tsai, F.Y.; Coruzzi, G.M.
#journal EMBO J. (1990) 9:323-332
#title Dark-induced and organ-specific expression of two asparagine
synthetase genes in Pisum sativum.
#cross-references WUID:90151604
#accession S11443
#molecule_type DNA
#residues 1-583 #label TSA
##cross-references EMBL:X52180; NID:g20651; PID:CA36430.1; PID:g20652

##cross-references EMBL:U18778; NID:g603592; PID:g603616; MIPS:YER024w
GENETICS
#map-position 5R
SUMMARY #length 923 #molecular-weight 103333 #checksum 767

Query Match 58.3%; Score 63; DB 2; Length 923;
Best Local Similarity 53.8%; Pred. NO. 4.41e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 722 MFMLSVLGEIRS 734
|||||:|||||:
QY 3 MFLLSRNTGEVRT 15

RESULT 9
ENTRY T08846 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
soybean
ALTERNATE_NAMES asparagine synthetase 2
ORGANISM #formal_name Glycine max #common_name soybean
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999

ACCESSIONS
REFERENCE T08846
#authors Hughes, C.A.; Beard, H.; Matthews, B.F.
#journal Plant Mol. Biol. (1997) 33:301-311
#title Molecular cloning and expression of two cDNAs encoding
asparagine synthetase in soybean.
#accession T08846
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-581 #label HUG
##cross-references EMBL:U77678; NID:g1778370
##experimental_source strain Century; leaf

GENETICS
#gene AS2
FUNCTION #description catalyzes ATP-dependent transfer of the amide group of
glutamine to aspartate generating asparagine and glutamate
#pathway asparagine biosynthesis
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 581 #molecular-weight 65609 #checksum 7265

Query Match 56.5%; Score 61; DB 2; Length 581;
Best Local Similarity 77.8%; Pred. NO. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMLMSRK 330
|||||:|||||:
QY 1 TPMLLSRN 9

RESULT 10
ENTRY AUPM2 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
garden pea
ALTERNATE_NAMES asparagine synthetase (glutamine-hydrolyzing)
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
18-Jun-1999

ACCESSIONS
REFERENCE S11443
#authors Tsai, F.Y.; Coruzzi, G.M.
#journal EMBO J. (1990) 9:323-332
#title Dark-induced and organ-specific expression of two asparagine
synthetase genes in Pisum sativum.
#cross-references WUID:90151604
#accession S11443
#molecule_type DNA
#residues 1-583 #label TSA
##cross-references EMBL:X52180; NID:g20651; PID:CA36430.1; PID:g20652

##cross-references EMBL:U18778; NID:g603592; PID:g603616; MIPS:YER024w
GENETICS
#map-position 5R
SUMMARY #length 923 #molecular-weight 103333 #checksum 767

Query Match 58.3%; Score 63; DB 2; Length 923;
Best Local Similarity 53.8%; Pred. NO. 4.41e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 722 MFMLSVLGEIRS 734
|||||:|||||:
QY 3 MFLLSRNTGEVRT 15

RESULT 9
ENTRY T08846 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
soybean
ALTERNATE_NAMES asparagine synthetase 2
ORGANISM #formal_name Glycine max #common_name soybean
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
11-Jun-1999

ACCESSIONS
REFERENCE T08846
#authors Hughes, C.A.; Beard, H.; Matthews, B.F.
#journal Plant Mol. Biol. (1997) 33:301-311
#title Molecular cloning and expression of two cDNAs encoding
asparagine synthetase in soybean.
#accession T08846
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-581 #label HUG
##cross-references EMBL:U77678; NID:g1778370
##experimental_source strain Century; leaf

GENETICS
#gene AS2
FUNCTION #description catalyzes ATP-dependent transfer of the amide group of
glutamine to aspartate generating asparagine and glutamate
#pathway asparagine biosynthesis
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 581 #molecular-weight 65609 #checksum 7265

Query Match 56.5%; Score 61; DB 2; Length 581;
Best Local Similarity 77.8%; Pred. NO. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMLMSRK 330
|||||:|||||:
QY 1 TPMLLSRN 9

RESULT 10
ENTRY AUPM2 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
garden pea
ALTERNATE_NAMES asparagine synthetase (glutamine-hydrolyzing)
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
18-Jun-1999

ACCESSIONS
REFERENCE S11443
#authors Tsai, F.Y.; Coruzzi, G.M.
#journal EMBO J. (1990) 9:323-332
#title Dark-induced and organ-specific expression of two asparagine
synthetase genes in Pisum sativum.
#cross-references WUID:90151604
#accession S11443
#molecule_type DNA
#residues 1-583 #label TSA
##cross-references EMBL:X52180; NID:g20651; PID:CA36430.1; PID:g20652

##residues 1-586 ##label MAT
##cross-references EMBL:X89409; NID:g897770; PIDN:CAA61589.1;
PID:g897771
##experimental_source strain B-129

GENETICS
#gene AS1
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 586 #molecular-weight 66461 #checksum 3372

Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFLMSRK 330
OY 1 TPFLMSRK 9

RESULT 13
ENTRY 569183 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
ORGANISM Lotus japonicus
#formal_name Lotus japonicus
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
ACCESSIONS 569183; S57932
REFERENCE 569182
#authors Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.;
#journal Clarkson, D.T.
#title Plant Mol. Biol. (1996) 30:883-897
Molecular cloning and characterization of asparagine
synthetase from Lotus japonicus: dynamics of asparagine
synthesis in N-sufficient conditions.
#cross-references MUID:96270368
#accession 569183
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-586 ##label MAT
#cross-references EMBL:X89410; NID:g897772; PIDN:CAA61590.1;
PID:g897773
##experimental_source strain B-129

GENETICS
#gene AS2
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 586 #molecular-weight 65969 #checksum 518

Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFLMSRK 330
OY 1 TPFLMSRK 9

RESULT 14
ENTRY AJPMN1 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 -
ALTERNATE_NAMES garden pea
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Jun-1999
ACCESSIONS S11444
REFERENCE S11443
#authors Tsai, F.Y.; Coruzzi, G.M.
#journal EMBO J. (1990) 9:323-332
#title Dark-induced and organ-specific expression of two asparagine
synthetase genes in Pisum sativum.
#cross-references MUID:90151604
#accession S11444

##molecule_type DNA
##residues 1-586 ##label TSA
##cross-references EMBL:X52179; NID:g20649; PIDN:CAA36429.1; PID:g20650
#comment This protein is one of a family of glutamine amidotransferases that
have dual specificity to utilize either glutamine or ammonia as a
substrate. Two functional domains have been identified: an
amino transfer domain, which catalyzes the ammonia-dependent reaction,
and a glutamine amide transfer domain, which binds glutamine and
transfers the amide to the amino transfer domain.

GENETICS
#gene AS1
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
FEATURE 175-586
#domain amino transfer
#active-site Cys (amide transfer) #status predicted
SUMMARY #length 586 #molecular-weight 66353 #checksum 3350

Query Match 56.5%; Score 61; DB 1; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFLMSRK 331
OY 1 TPFLMSRK 9

RESULT 15
ENTRY T02978 #type complete
TITLE asparagine synthase (EC 6.3.5.4) - maize
ALTERNATE_NAMES asparagine synthetase
ORGANISM #formal_name Zea mays #common_name maize
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
ACCESSIONS T02978
REFERENCE 214805
#authors Chevalier, C.; Bourgeois, E.; Just, D.; Raymond, P.
#journal Plant J. (1996) 9:1-11
#title Metabolic regulation of asparagine synthetase gene expression
in maize (Zea mays L.) root tips.
#cross-references MUID:96158342
#accession T02978
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-586 ##label CHE
##cross-references EMBL:X82849; NID:g984261; PIDN:CAA58052.1;
PID:g984262
##experimental_source cultivar var.DEA; root meristem

GENETICS
#gene AS
CLASSIFICATION #superfamily asparagine synthase (glutamine-hydrolyzing)
KEYWORDS asparagine biosynthesis; ligase
SUMMARY #length 586 #molecular-weight 66577 #checksum 5815

Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.10e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFLMSRK 330
OY 1 TPFLMSRK 9

Search completed: Sat May 13 07:16:22 2000
Job time : 8 secs.


```
KW Calcium-binding; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 999 DESMOGLEIN 3.
FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 616 640 POTENTIAL.
FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 268 CADHERIN 2.
FT REPEAT 269 383 CADHERIN 3.
FT REPEAT 386 499 CADHERIN 4.
FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 459 459 POTENTIAL.
FT CARBOHYD 545 545 POTENTIAL.
SQ SEQUENCE 999 AA; 107503 MM; 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 108; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 4.18e-13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPFLLSRNTGEVRT 220
OY 1 TPFLLSRNTGEVRT 15

RESULT 2
ID DSG1_HUMAN STANDARD: PRT; 1049 AA.
AC 002413;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).
GN DSG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KERATINOCYTES;
RX Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
  Arreman J., Rutman A.J., Pidsley S.C., Watt F.M., Rees D.A.,
  Buxton R.S., Magee A.I.;
  "Desmosomal glycoprotein DGL, a component of intercellular desmosome
  junctions, is related to the cadherin family of cell adhesion
  molecules."
RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
RL -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME
  INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
  FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCULIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
  (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; X56654; CA39976.1; -
DR PIR; S16906; IJHUG1.
DR HSSP; P09803; 1EDH.
DR MIM; 125670; -
DR PROSITE; PS00232; CADHERIN. 2.
DR PFM; PF00028; cadherin. 4.
```

```
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 1049 DESMOGLEIN 1.
FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 546 570 POTENTIAL.
FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 270 CADHERIN 2.
FT REPEAT 271 385 CADHERIN 3.
FT REPEAT 386 497 CADHERIN 4.
FT REPEAT 813 839 DESMOGLEIN REPEAT 1.
FT REPEAT 840 869 DESMOGLEIN REPEAT 2.
FT REPEAT 870 899 DESMOGLEIN REPEAT 3.
FT REPEAT 900 927 DESMOGLEIN REPEAT 4.
FT REPEAT 928 956 DESMOGLEIN REPEAT 5.
FT DOMAIN 969 1019 GLY/SER-RICH.
FT CARBOHYD 36 36 POTENTIAL.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
SQ SEQUENCE 1049 AA; 113715 MM; EEE125655B9D6619 CRC64;

Query Match 81.5%; Score 88; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 1.09e-07;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFLLSRNTGEVRT 220
OY 1 TPFLLSRNTGEVRT 15
```

```
RESULT 3
ID DSG1_BOVIN STANDARD: PRT; 1043 AA.
AC 003763;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).
GN DSG1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RX Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
  Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RL [2]
RL SEQUENCE OF 44-1043 FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RX MEDLINE; 9168965.
RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
  Zimbelmann R., Franke W.W.;
  "Identification of desmoglein, a constitutive desmosomal
  glycoprotein, as a member of the cadherin family of cell adhesion
  molecules."
RT Eur. J. Cell Biol. 53:1-12(1990).
RL [3]
RL REVISIONS, AND SEQUENCE OF 101-123.
RX MEDLINE; 92037656.
RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
  Franke W.W.;
  "Complete amino acid sequence of the epidermal desmoglein precursor
  polypeptide and identification of a second type of desmoglein gene."
RT Eur. J. Cell Biol. 55:200-208(1991).
RL [4]
RP SEQUENCE OF 44-493 FROM N.A.
RX MEDLINE; 91097553.
RA Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;
  "Desmoglein shows extensive homology to the cadherin family of cell
  adhesion molecules."
RT
```

```

RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58466; CAA41380.1; -
DR EMBL: X57784; CAA40930.1; -
DR EMBL: M58165; AAA62709.1; -
DR PIR: S14603; IJBOG1.
DR HSSP: P09803; IEDH.
DR PROSITE: P500232; CADHERIN; 2.
DR PFAM: PF00028; cadherin; 3.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1043
FT DOMAIN 50 548
FT TRANSMEM 549 573
FT DOMAIN 574 1043
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 498
FT REPEAT 499 845
FT REPEAT 846 875
FT REPEAT 876 905
FT REPEAT 906 933
FT REPEAT 934 962
FT DOMAIN 963 1012
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 486 486
FT CONFLICT 124 124
SQ SEQUENCE 1043 AA; 112243 MW; ADE46133F8B7C11 CRC64;

Query Match
Best Local Similarity 75.0%; Score 81; DB 1; Length 1043;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 206 SPMEIINRTGEIRT 220
OY 1 TPMEFLSRNTGEVRT 15

RESULT 4
ID ASN2_YEAST STANDARD: PRT: 571 AA.
AC P49090.
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 01-NOV-1997 (Rel. 35; Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
GN ASN2 OR YGR124W OR G6356.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;

```

```

RX MEDLINE: 97197982.
RA van Dyck L., Tettelin H., Purnelle B., Goffeau A.;
RT An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT unknown open reading frames; the gene for an Asn synthase, remnants
RT of Ty and three tRNA genes.
RT yeast 13:171-176(1997).
RP [2]
RP SEQUENCE FROM N.A.
RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla E.,
RA Nawrocki A., del Bino S., Goffeau A.;
RA Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83099; CAA58159.1; -
DR EMBL: 272909; CAA97135.1; -
DR SGD: L0003156; ASN2.
DR PROSITE: P500443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KW Init_Mer
FT ACT_SITE 0 0
FT ACT_SITE 1 1
FT ACT_SITE 1 1
SQ SEQUENCE 571 AA; 64461 MW; 086D03EC77E6F00A CRC64;

Query Match
Best Local Similarity 88.9%; Score 65; DB 1; Length 571;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 345 TPMEFLSRK 353
OY 1 TPMEFLSRN 9

RESULT 5
ID ASN1_YEAST STANDARD: PRT: 571 AA.
AC P49089.
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 15-FEB-2000 (Rel. 39; Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 1 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).
GN ASN1 OR YPR145W OR P9659.3
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA Dang V.D., Valens M., Bolotin-Fukuhara M., Dalsman-Fornier B.;
RA "Cloning of the ASN1 and ASN2 genes encoding asparagine synthetases
RA in Saccharomyces cerevisiae: differential regulation by the
RA CCAAT-box-binding factor."
RL Mol. Microbiol. 22:681-692(1996).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RC Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,

```


CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52180; CAA36430.1; -
CC EMBL: Y13322; CAA73763.1; -
CC PIR: S11443; ASPMN2.
CC PROSITE: PS00443; GATASE_TYPE_II; 1.
CC PFAM: PF00310; GATase_2; 1.
CC PFM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; Glutamine amidotransferase;
CC Multigene family.
CC INIT_MER 0 0 BY SIMILARITY.
CC ACT_SITE 1 1 GATASE (BY SIMILARITY).
CC CONFLICT 33 33 E -> D (IN REF. 2).
CC CONFLICT 44 44 Y -> F (IN REF. 2).
CC CONFLICT 47 47 Q -> H (IN REF. 2).
CC CONFLICT 77 77 I -> L (IN REF. 2).
CC SEQUENCE 582 AA; 65518 MW; F98546B4B3456265 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 582;
Best Local Similarity 77.8%; Pred. No. 3.25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 10
ID ASN_ARATH STANDARD; PRT; 583 AA.
AC P49078;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE DEPENDENT ASPARAGINE SYNTHETASE).
GN ASN1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsids.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA;
RX MEDLINE: 95148732.
RA Lam H.M., Peng S.S., Coruzzi G.M.;
RT "Metabolic regulation of the gene encoding glutamine-dependent
RT asparagine synthetase in Arabidopsis thaliana."
RT Plant Physiol. 106:1347-1357(1994).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE - AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: L29083; AAA74359.1; -
DR HSSP: P17169; IGMS.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFM: PF00310; GATase_2; 1.
DR PFM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; Glutamine amidotransferase;
CC Multigene family.
CC INIT_MER 0 0 BY SIMILARITY.
CC ACT_SITE 1 1 GATASE (BY SIMILARITY).
CC SEQUENCE 583 AA; 65489 MW; C2C257781D8C7D23 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 583;
Best Local Similarity 77.8%; Pred. No. 3.25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
RESULT 11
ID ASN_TRIVS STANDARD; PRT; 585 AA.
AC O24661;
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-FEB-2000 (rel. 39, Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE DEPENDENT ASPARAGINE SYNTHETASE).
GN AS.
OS Triphysaria versicolor.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids I; Lamiales; Scrophulariaceae;
OC Triphysaria.
OC [1]
RN SEQUENCE FROM N.A.
RP TISSUE-ROOT;
RX MEDLINE: 99051324.
RA Delavault P., Strabrook E., Albrecht H., Wrobel R., Yoder J.I.;
RT "Host-root exudates increase gene expression of asparagine synthetase
RT in the roots of a hemiparasitic plant triphysaria versicolor
RT (Scrophulariaceae)."
RL Gene 222:155-162(1998).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE - AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF014055; AAD5033.1; -
DR EMBL: AF014056; AAD5034.1; -
DR EMBL: AF014057; AAD5035.1; -
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFM: PF00310; GATase_2; 1.
DR PFM: PF00733; Asn_synthase; 1.
CC LIGase: Asparagine biosynthesis; Glutamine amidotransferase;
CC Multigene family.
CC INIT_MER 0 0 BY SIMILARITY.
CC ACT_SITE 1 1 GATASE (BY SIMILARITY).
CC SEQUENCE 585 AA; 65560 MW; FAF2E622912117B5 CRC64;
DB 321 TPWFLLSRK 329
OY 1 TPWFLLSRN 9
Query Match 56.5%; Score 61; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 3.25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPEFLMSRK 329
 QY 1 TPEFLMSRN 9

RESULT 12
 ID ASN2 LOTJA STANDARD: PRT: 585 AA.
 AC P49093:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
 GN AS2.
 OS Lotus Japonicus.
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 OC core eudicots: Rosidae: eurosids I: Fabales; Fabaceae; Papilionoideae;
 OC Lotus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. GIFU / B-129;
 RX MEDLINE: 96370368.
 RA Waterhouse R.N., Smyth A.J., Massoneau A., Prosser I.M.,
 RA Clarkson D.T.;
 RT "Molecular cloning and characterisation of asparagine synthetase from
 RT Lotus japonicus: dynamics of asparagine synthesis in N-sufficient
 RT conditions.";
 RT Plant Mol. Biol. 30:883-897(1996).
 RL
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X89410; CAA61590.1; -
 CC DR HSSP: P17169; IGMS.
 CC DR PROSITE: PS00443; GATASE_TYPE_II; 1.
 CC DR PFAM: PF00310; GATase_2; 1.
 CC DR PFAM: PF00733; Asn_synthase; 1.
 CC KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
 CC Multigene family.
 CC FT INIT_MET 0 BY SIMILARITY.
 CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
 CC SQ SEQUENCE 585 AA; 65838 MW; F56DCA015F73451 CRC64;

Query Match 56.5%: Score 61; DB 1; Length 585;
 Best Local Similarity 77.8%: Pred. No. 3.25e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPEFLMSRK 329
 QY 1 TPEFLMSRN 9

RESULT 13
 ID ASN1 PEA STANDARD: PRT: 585 AA.
 AC P19251; O49925;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ASPARAGINE SYNTHETASE, NODULE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
 GN AS1.

OS Pisum sativum (Garden pea).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons;
 OC core eudicots: Rosidae: eurosids I: Fabales; Fabaceae; Papilionoideae;
 OC Pisum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SPARKLE; TISSUE=ROOT NODULES;
 RX MEDLINE: 90151604.
 RA Tsai F.Y., Coruzzi G.M.;
 RT "Dark-induced and organ-specific expression of two asparagine
 RT synthetase genes in Pisum sativum.";
 RL EMOB J. 9:323-332(1990).
 RN [2]
 RP SEQUENCE OF 1-83 FROM N.A.
 RC STRAIN-CV. FELTHAM FIRST;
 RA Ngai N., Tsai F.Y., Coruzzi G.M.;
 RT "Light-induced transcriptional repression of the pea As1 gene:
 RT identification of cis-elements and transactors.";
 RL Submitted (MAY 1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: to the EMBL/GenBank/DBJ databases.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- TISSUE SPECIFICITY: NODULE.
 CC -1- INDUCTION: DARK-INDUCED.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X52179; CAA36429.1; -
 CC DR EMBL: Y13321; CAA37362.1; -
 CC DR PIR: S11444; AJPANI.
 CC DR HSSP: P17169; IGMS.
 CC DR PROSITE: PS00443; GATASE_TYPE_II; 1.
 CC DR PFAM: PF00310; GATase_2; 1.
 CC DR PFAM: PF00733; Asn_synthase; 1.
 CC KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
 CC Multigene family.
 CC FT INIT_MET 0 BY SIMILARITY.
 CC FT ACT_SITE 1 GATASE (BY SIMILARITY).
 CC SQ SEQUENCE 585 AA; 66222 MW; 6CEBC4417B120ECD CRC64;

Query Match 56.5%: Score 61; DB 1; Length 585;
 Best Local Similarity 77.8%: Pred. No. 3.25e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPEFLMSRK 330
 QY 1 TPEFLMSRN 9

RESULT 14
 ID ASNS MAIZE STANDARD: PRT: 585 AA.
 AC P49094;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
 DE DEPENDENT ASPARAGINE SYNTHETASE).
 GN ASN1 OR AS.
 OS Zea mays (Maize).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida; Poales;
 OC Poaceae; Zea.
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. DEA: TISSUE-ROOT MERISTEM:
RA MEDLINE: 96158342.
RX Chevalier C., Bourgeois E., Just D., Raymond P.;
RT "Metabolic regulation of asparagine synthetase gene expression in
RL maize (Zea mays L.) root tips.";
RL Plant J. 9:1-11(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X82849; CA58052.1;
DR MAZEDB: 79071;
DR PROSITE: PS00443; GATASE_TYPE_II: 1.
DR PFAM: PF00310; GATase_2: 1.
DR PFAM: PF00733; Asn_synthase: 1.
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA; 66446 MM; 095CE2F99973797E CRC64;

Query Match 56.5%; Score 61; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 3.25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPMFLMSRK 329
QY 1 TPMFLMSRN 9

RESULT 15
ID ASNS_BRAOL STANDARD: PRT; 585 AA.
AC P49091;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-1998 (Rel. 36, Last annotation update)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-
DE DEPENDENT ASPARAGINE SYNTHETASE).
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Brassica.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. SHOGUN;
RC Downs C.G., Pogson B.J., Davies K.M., Almira E.C.;
RT "An asparagine synthetase cDNA clone from Broccoli (Brassica oleracea
RT L.)".
RL (in) Plant Gene Register PGR95-016.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X84448; CA59138.1;
DR PROSITE: PS00443; GATASE_TYPE_II: 1.
DR PFAM: PF00310; GATase_2: 1.
DR PFAM: PF00733; Asn_synthase: 1.
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA; 65541 MM; A73083B98A22FE6 CRC64;

Query Match 56.5%; Score 61; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 3.25e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMFLMSRK 330
QY 1 TPMFLMSRN 9

Search completed: Sat May 13 07:12:43 2000
Job time : 90 secs.

```

 WIREIMAGE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MSPRCH_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:13:02 2000; MasPar time 158.22 Seconds

Tabular output not generated. 6.573 Million cell updates/sec

Title: >US-08-991-628-4

Description: (1-15) from US08991628.pap

Perfect Score: 108

Sequence: 1 TFMFLSRNTGEVRT 15

Scoring table: PAM 150

Searched: 225878 segs, 69334122 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database:

sptrembl12
 1:sp_invertebrate 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_protent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.384; Variance 31.843; scale 0.860

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	90	83.3	993	11	035902	DESMOGLEIN 3 (FRAGMENT 4.79e-08
2	67	62.0	878	4	09Y5C2	PROTODACHERIN GAMMA C5 2.77e-02
3	67	62.0	944	4	09Y5F6	PROTODACHERIN GAMMA C5 2.77e-02
4	64	59.3	336	5	009661	HYPOTHETICAL 37.7 KD P 1.34e-01
5	63	58.3	291	2	004824	ORF6 PROTEIN PRECURSOR 2.24e-01
6	62	57.4	643	5	018290	HYPOTHETICAL PROTEIN C 3.72e-01
7	61	56.5	557	3	042902	ASPARAGINE SYNTHETASE. 6.17e-01
8	61	56.5	578	10	0925T7	ASPARAGINE SYNTHETASE. 6.17e-01
9	61	56.5	578	10	0925T6	ASPARAGINE SYNTHETASE. 6.17e-01
10	61	56.5	581	10	0925T6	ASPARAGINE SYNTHETASE. 6.17e-01
11	61	56.5	584	10	0925T3	ASPARAGINE SYNTHETASE. 6.17e-01
12	61	56.5	585	10	065329	ASPARAGINE SYNTHETASE. 6.17e-01
13	61	56.5	586	10	093618	ASPARAGINE SYNTHETASE. 6.17e-01
14	61	56.5	586	10	040328	ASPARAGINE SYNTHETASE. 6.17e-01
15	61	56.5	586	10	024483	ASPARAGINE SYNTHETASE. 6.17e-01
16	61	56.5	871	4	09Y5C7	PROTODACHERIN GAMMA C4 6.17e-01
17	61	56.5	938	4	09Y5F7	PROTODACHERIN GAMMA C4 6.17e-01
18	59	54.6	148	1	058871	148AA LONG HYPOTHETICAL 1.66e+00
19	59	54.6	820	1	015039	KIAA0327 PROTEIN. 1.66e+00
20	59	54.6	852	1	09YDL5	852AA LONG HYPOTHETICAL 1.66e+00

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	035902	90	83.3%	993	11	035902	DESMOGLEIN 3 (FRAGMENT 4.79e-08
2	09Y5C2	67	62.0%	878	4	09Y5C2	PROTODACHERIN GAMMA C5 2.77e-02
3	09Y5F6	67	62.0%	944	4	09Y5F6	PROTODACHERIN GAMMA C5 2.77e-02
4	009661	64	59.3%	336	5	009661	HYPOTHETICAL 37.7 KD P 1.34e-01
5	004824	63	58.3%	291	2	004824	ORF6 PROTEIN PRECURSOR 2.24e-01
6	018290	62	57.4%	643	5	018290	HYPOTHETICAL PROTEIN C 3.72e-01
7	042902	61	56.5%	557	3	042902	ASPARAGINE SYNTHETASE. 6.17e-01
8	0925T7	61	56.5%	578	10	0925T7	ASPARAGINE SYNTHETASE. 6.17e-01
9	0925T6	61	56.5%	578	10	0925T6	ASPARAGINE SYNTHETASE. 6.17e-01
10	0925T6	61	56.5%	581	10	0925T6	ASPARAGINE SYNTHETASE. 6.17e-01
11	0925T3	61	56.5%	584	10	0925T3	ASPARAGINE SYNTHETASE. 6.17e-01
12	065329	61	56.5%	585	10	065329	ASPARAGINE SYNTHETASE. 6.17e-01
13	093618	61	56.5%	586	10	093618	ASPARAGINE SYNTHETASE. 6.17e-01
14	040328	61	56.5%	586	10	040328	ASPARAGINE SYNTHETASE. 6.17e-01
15	024483	61	56.5%	586	10	024483	ASPARAGINE SYNTHETASE. 6.17e-01
16	09Y5C7	61	56.5%	871	4	09Y5C7	PROTODACHERIN GAMMA C4 6.17e-01
17	09Y5F7	61	56.5%	938	4	09Y5F7	PROTODACHERIN GAMMA C4 6.17e-01
18	058871	59	54.6%	148	1	058871	148AA LONG HYPOTHETICAL 1.66e+00
19	015039	59	54.6%	820	1	015039	KIAA0327 PROTEIN. 1.66e+00
20	09YDL5	59	54.6%	852	1	09YDL5	852AA LONG HYPOTHETICAL 1.66e+00

```

OC Eutheria: Primates: Catarrhini: Hominoidea: Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99308636.
RA WU Q., MANIATIS T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes."
RL Cell 97:779-790(1999).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF152526; AAD43786.1; -.
DR PROSITE: PS00232; CADHERIN; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 878 AA: 95192 MW: 078070CA CRC32:

Query Match
Best Local Similarity 62.0%; Score 67; DB 4; Length 878;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 618 LFLVSTHTGEVRT 630
OY 3 MFLSRTTGEVRT 15

RESULT 3
ID 09Y5F6 PRELIMINARY: PRT: 944 AA.
AC 09Y5F6:
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DE 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
DE PROTOCADHERIN GAMMA C5.
GN PCDH-GAMMA-C5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN:
RX MEDLINE: 99308636.
RA WU Q., MANIATIS T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes."
RL Cell 97:779-790(1999).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF152526; AAD43786.1; -.
DR PROSITE: PS00232; CADHERIN; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 944 AA: 101920 MW: 2692C1D3 CRC32:

Query Match
Best Local Similarity 62.0%; Score 67; DB 4; Length 944;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 618 LFLVSTHTGEVRT 630
OY 3 MFLSRTTGEVRT 15

RESULT 4
ID 009661 PRELIMINARY: PRT: 336 AA.
AC 009661:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE HYPOTHEICAL 37.7 KD PROTEIN ZK177.6 IN CHROMOSOME II.
GN ZK177.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2:
RA ANDERSON K.;
SQ Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- SIMILARITY: WEAK, TO YEAST PROTEIN CDC20.
DR EMBL: U21321; AAB36970.1; -.
DR WORMPEP: ZK177.6; CE02095.
DR PFAM: PF00400; WD40; 1.
KW Hypothetical protein.
SQ SEQUENCE 336 AA: 37668 MW: 2DC77826 CRC32:

Query Match
Best Local Similarity 59.3%; Score 64; DB 5; Length 336;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 187 PMFTIDEHTGEVRT 200
OY 2 PMFLSRTTGEVRT 15

RESULT 5
ID 004824 PRELIMINARY: PRT: 291 AA.
AC 004824:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-JAN-1999 (TRENBLREL. 09, Last annotation update)
DE ORF6 PROTEIN PRECURSOR.
OS Salmonella typhimurium.
OC plasmid 90 kb virulence.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2:
RX MEDLINE: 93316852.
RA FRIEDRICH M.J., KINSEY N.E., VILA J., KADNER R.J.;
RT "Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence
RT plasmid of Salmonella typhimurium: the presence of fimbrial
RT biosynthetic genes."
RL MOL. Microbiol. 8:543-558(1993).
DR EMBL: L08613; AAC36963.1; -.
KW Plasmid; Signal.
FT SIGNAL 1 23
FT CHAIN 24 291 ORF6 PROTEIN.
SQ SEQUENCE 291 AA: 30683 MW: C4D5ADDF CRC32:

Query Match
Best Local Similarity 58.3%; Score 63; DB 2; Length 291;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 128 TLMFSLR-TGEVRT 141
OY 1 TPEFLSRTTGEVRT 15

RESULT 6
ID 018290 PRELIMINARY: PRT: 643 AA.
AC 018290: Q20732.
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
DE 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE HYPOTHEICAL PROTEIN C29A12.6.
GN C29A12.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA WILKINSON J.;
SQ Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z73970; CAA98248.1; -.
DR EMBL: Z77664; CAA98248.1; JOINED.
DR EMBL: Z77664; CAB01217.1; JOINED.
DR EMBL: Z73970; CAB01217.1; JOINED.
SQ SEQUENCE 643 AA: 73571 MW: C0092D81 CRC32:

Query Match
Best Local Similarity 57.4%; Score 62; DB 5; Length 643;

```



```

ID 092523 PRELIMINARY: PRT: 584 AA.
AC 092523:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE TYPE II (EC 6.3.5.4).
GN AS2.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euhypophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Phaseolus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. GREAT NORTHERN; TISSUE-ROOT;
RA OSUNA D., GALVEZ G., PINEDA M., AGUILAR M.;
RT "R-PCR cloning and expression of a cDNA encoding a type-II asparagine
RT synthetase."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ009952; CA08913.1; -.
DR HSSP: P17169; IGMS.
DR MENDEL: 34839; Phvu:1042;34839.
DR LIGase.
KW SEQUENCE 584 AA; 65743 MW; CDE20571 CRC32;

Query Match 56.5%; Score 61; DB 10; Length 584;
Best Local Similarity 77.8%; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMFLMSRK 330
Qy 1 TPMFLMSRN 9

RESULT 12
ID 065329 PRELIMINARY: PRT: 585 AA.
AC 065329:
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE.
GN AS.
OS Eleagnus umbellata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euhypophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Rosales; Elaeagnaceae; Elaeagnus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ROOT NODELE;
RA KIM H.-B., AN C.-S.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061740; AAC16325.1; -.
DR MENDEL: 28685; elaum:1042;29685.
DR PFAM: PF00733; Asn_synthase; 1.
DR PFAM: PF00310; GATase_2; 1.
DR SEQUENCE 585 AA; 65878 MW; 8A12777D CRC32;

Query Match 56.5%; Score 61; DB 10; Length 585;
Best Local Similarity 77.8%; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPMFLMSRK 330
Qy 1 TPMFLMSRN 9

RESULT 13
ID P93618 PRELIMINARY: PRT: 586 AA.
AC P93618:
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE (EC 6.3.5.4) (ASPARAGINE SYNTHASE

```

```

DE (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
DE (GLUTAMINE-HYDROLYSING)).
GN VFAS1.
OS Vicia faba (broad bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euhypophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Vicia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KLEINE THUNDERINGER; TISSUE-ROOT NODELE;
RA KUSTER H., ALBUS U., FRUEHLING M., TCHETKOVA S.A., TIKHONOVITCH I.A.,
RA PUEHLER A., PERLICK A.M.;
RL Plant Sci. 124:89-95(1977).
CC -1 CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
DR EMBL: Z72354; CA96526.1; -.
DR HSSP: P17169; IGMS.
DR MENDEL: 11023; Vlcfa:1042;11023.
DR PFAM: PF00733; Asn_synthase; 1.
DR PFAM: PF00310; GATase_2; 1.
DR LIGase.
KW SEQUENCE 586 AA; 66236 MW; AD8BD927 CRC32;

Query Match 56.5%; Score 61; DB 10; Length 586;
Best Local Similarity 77.8%; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPMFLMSRK 331
Qy 1 TPMFLMSRN 9

RESULT 14
ID 040328 PRELIMINARY: PRT: 586 AA.
AC 040328:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE.
GN Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euhypophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Medicago.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SARANAC;
RA GANTT S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: LA0327; AAB4058.1; -.
DR HSSP: P17169; IGMS.
DR MENDEL: 9036; Medsa:1042;9036.
DR PFAM: PF00733; Asn_synthase; 1.
DR PFAM: PF00310; GATase_2; 1.
DR SEQUENCE 586 AA; 66462 MW; 962C3A3F CRC32;

Query Match 56.5%; Score 61; DB 10; Length 586;
Best Local Similarity 77.8%; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPMFLMSRK 331
Qy 1 TPMFLMSRN 9

RESULT 15
ID 024483 PRELIMINARY: PRT: 586 AA.
AC 024483:
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE ASPARAGINE SYNTHETASE.

```

OS Medicago sativa (Alfalfa).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae:
OC Medicago.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97432147.
RA SHI L., TWARY S.N., YOSHIOKA H., GREGGSON R.G., MILLER S.S.,
RA SAMAC D.A., GANTT J.S., UNKEFER P.J., VANCE C.P.:
RT "Nitrogen assimilation in alfalfa: Isolation and characterization of
RT an asparagine synthetase gene showing enhanced expression in root
RT nodules and dark-adapted leaves.";
RL Plant Cell 9:1339-1356(1997).
DR EMBL: U89923; AAB81011.1; -.
DR HSSP: P17169; IGMS.
DR MENDEL: 27006; Medsa:1042;27006.
DR PFAM: PF00733; Asn-synthase; 1.
DR PFAM: PF00310; GATase-2; 1.
SO SEQUENCE 586 AA; 66448 MW; 63C8B6BA CRC32;

Query Match 56.58; Score 61; DB 10; Length 586;
Best Local Similarity 77.88; Pred. No. 6.17e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 323 TPWFILSRK 331
QY 1 TPWFILSRN 9

Search completed: Sat May 13 07:15:54 2000
Job time : 172 secs.

THIS PAGE BLANK (USPTO)

US-08-991-628-5a.rag

Mon May 15 08:08:55 2000

RESULT 2
 R30742 standard: Protein: 999 AA.
 R30742: (first entry)
 14-JUN-1993 (first entry) 130kd antigen.
 Human pemphigus vulgaris 130kd antigen; autoantibodies:
 E pemphigus vulgaris: skin disease; glycoprotein: cell adhesion.
 W keratinocyte cell surface antigen.
 N Homo sapiens.
 US7798918-A.
 15-DEC-1992.
 27-NOV-1991: 798918.
 (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PA Amagel M, Klaus-Kovtun V, Stanley JR.
 PI WPI: 93-067436/08.
 DR N-PSDB: Q35992.
 PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
 PS diagnostic and therapeutic uses
 CC disclosure: Fig 7: 50pp: English.
 CC This sequence is the pemphigus vulgaris 130kd antigen. The protein
 CC and its encoding DNA may be used in the diagnosis and treatment of
 CC pemphigus vulgaris. It is thought that the antigen may be a cell
 CC adhesion molecule.
 CC Sequence 999 AA:
 SQ

Query Match 100.0% Score 113; DB 1; Length 999;
 Best Local Similarity 100.0% Pred. No. 3,35e-05; Indels 0; Gaps 0;
 Matches 15; Conservative

Db 251 CECNIKVDVNDNFP 265
 1 CECNIKVDVNDNFP 15

Query Match 100.0% Score 113; DB 1; Length 999;
 Best Local Similarity 100.0% Pred. No. 3,35e-05; Indels 0; Gaps 0;
 Matches 15; Conservative

RESULT 3
 W07908 standard: Protein: 614 AA.
 AC W07908:
 29-JAN-1997 (first entry) protein extracellular region;
 DE pemphigus vulgaris antigen; fusion protein; diagnosis;
 KM Autoantibody: immunoglobulin G; IgG1; blister; skin disease;
 KM treatment: pemphigus vulgaris; PV; bulla; blister; skin disease;
 KM dermatology.
 OS Homo sapiens.
 PN J08188540-A.
 23-JUL-1996.
 30-JUN-1995: 165632.
 30-JUN-1994: JP-173291.
 PR (NISH/) NISHIKAWA T.
 PA WPI: 96-388562/39.
 DR Fused protein recognised by pemphigus vulgaris auto-antibody -
 PT useful to treat and diagnose pemphigus vulgaris
 PS Claim 1: Page 7-9: 9pp: Japanese.
 CC W07908 represents the human pemphigus vulgaris (PV) antigen
 CC extracellular region. The PV antigen is produced in patients with
 CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
 CC relapsing disease causing suppurative, intra-epidermal bullae
 CC (vesicles) of the skin and mucous membranes, which is fatal if
 CC untreated. The PV antigen was fused to a human IgG1 hinge region
 CC and the resulting fusion protein is useful to treat or diagnose
 CC pemphigus vulgaris.
 CC Sequence 614 AA:
 SQ

Query Match 90.3% Score 102; DB 1; Length 614;
 Best Local Similarity 93.3% Pred. No. 6,39e-04; Indels 0; Gaps 0;
 Matches 14; Conservative

Db 250 CECNIKVDVNDNFP 264
 1 CECNIKVDVNDNFP 15

RESULT 4

RESULT 5
 R86865 standard: Protein: 787 AA.
 AC R86865:
 27-AUG-1996 (first entry)
 DE Human protocadherin pc3.
 KM Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KM catenin; therapy.
 OS Homo sapiens.
 PN M09600289-A1.
 04-JAN-1996.
 26-JUN-1995: U08071.
 27-JUN-1994: US-268161.
 PR (DOHE-) DOHENY EYE INST.
 PA Suzuki S.
 PI WPI: 96-068873/07.
 DR N-PSDB: T03572.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PS Claim 1: Page 115-119: 14pp: English.
 CC R86865-R86867 represent the sequences for three protocadherins. This
 CC sequence represents the human protocadherin pc3. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherins adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to

Query Match 78.8% Score 89; DB 1; Length 778;
 Best Local Similarity 80.0% Pred. No. 1,95e-02; Indels 0; Gaps 0;
 Matches 12; Conservative

Db 253 CECNIKVDVNDNFP 267
 1 CECNIKVDVNDNFP 15

Query Match 78.8% Score 89; DB 1; Length 778;
 Best Local Similarity 80.0% Pred. No. 1,95e-02; Indels 0; Gaps 0;
 Matches 12; Conservative

RESULT 6
 W15489 standard: Protein: 778 AA.
 AC W15489:
 17-JUN-1997 (first entry)
 DE pemphigus foliaceus antigen-IgG constant region fusion protein.
 DE pemphigus foliaceus; autoantibody; constant region; IgG;
 KM extracellular region; antigen; hinge portion; skin;
 KM dermatitis herpetiformis; fusion protein; detection; ss.
 OS Chimeric - Homo sapiens.
 FT Key Location/Qualifiers
 1..345
 /note="pemphigus foliaceus antigen protein"
 FT domain
 FT J09077800-A.
 25-MAR-1997.
 12-SEP-1995: 260899.
 12-SEP-1995: JP-260899.
 PR (NISH/) NISHIKAWA T.
 PA WPI: 97-241758/22.
 DR P-PSDB: T66428.
 DR Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
 PT through the hinge region used to treat pemphigus foliaceus
 PS Claim 1: Page 10-12: 17pp: Japanese.
 CC This sequence represents a fused protein recognised by pemphigus
 CC foliaceus patient autoantibody which comprises the constant region
 CC of IgG linked to the extracellular region of pemphigus foliaceus is
 CC antigen protein through the hinge portion. Pemphigus foliaceus
 CC a chronic, generalised, vesicular and scaling skin eruption similar
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen is
 CC protein is useful to treat pemphigus foliaceus. The antigen is
 CC especially administered through an adsorbent upon which the fusion
 CC protein is immobilised via a carrier. The fusion protein is useful
 CC useful for detecting pemphigus foliaceus antibodies which is useful
 CC in immunodiagnosis. The fusion protein has little or no side effects.
 CC Sequence 778 AA:
 SQ

Query Match 78.8% Score 89; DB 1; Length 778;
 Best Local Similarity 80.0% Pred. No. 1,95e-02; Indels 0; Gaps 0;
 Matches 12; Conservative

Db 253 CECNIKVDVNDNFP 267
 1 CECNIKVDVNDNFP 15

Query Match 78.8% Score 89; DB 1; Length 778;
 Best Local Similarity 80.0% Pred. No. 1,95e-02; Indels 0; Gaps 0;
 Matches 12; Conservative

CC proteins also allows investigation of the structure and function of
 CC cadherin. The cadherin proteins may be isolated by using anti-cadherin
 CC antibodies. These antibodies may also be used to modulate the activity
 CC of cadherin and to determine the tissue specific distribution of cadherin
 CC proteins. Each subclass of cadherins has a unique tissue distribution
 CC pattern.
 SQ Sequence 780 AA;

Query Match 50.4%; Score 57; DB 1; Length 780;
 Best Local Similarity 81.8%; Pred. No. 5.45e+01;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 135 IKVNDVNDNMP 145
 ||| |||||
 QY 5 IKVNDVNDNMP 15

RESULT 9
 ID W84311 standard; Protein: 83 AA.

AC W84311;
 DT 18-MAR-1999 (first entry)
 DE Oryza sativa sequence used to exemplify the invention.
 KW DNA sequence comparison.

OS Oryza sativa.
 PN J10334104-A.
 PD 18-DEC-1998.
 PF 05-MAR-1998; 053102.
 PR 31-MAR-1997; JP-079586.
 PA (HITA) HITACHI LTD.
 DR WPI: 99-110965/10.
 PT Comparison of DNA base sequences - using accurate observations of
 PT insertions and deletions

PS Disclosure: Page 16-17; 24pp; Japanese.

CC The present sequence is used to demonstrate the method of the invention.
 CC The specification describes the comparison of DNA base sequences. The
 CC method involves 3 steps of: (1) division of a first and second nucleic
 CC acid sequence into base groups of three base lengths, and translation
 CC into amino acids; (2) all base and amino acid insertions and deletions
 CC of the two nucleic acid sequences are observed and the amino acid
 CC sequences encoded by these nucleic acid sequences are compared; and
 CC (3) adjacent bases and amino acids are compared and step (1) repeated
 CC after shifting the translation frame by one base towards the 3' end or
 CC shifting the translation frame one base towards the 3' end, but
 CC encompassing four bases, and not translating the second or third base.
 CC The new method is useful for comparing DNA base sequences.

SQ Sequence 83 AA;

Query Match 49.6%; Score 56; DB 1; Length 83;
 Best Local Similarity 40.0%; Pred. No. 6.88e+01;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 68 CESSIYOYIDEARF 82
 ||:|:|:|:|:|:|
 QY 1 CECNIKVKVDNDFP 15

RESULT 10
 ID R55273 standard; Protein: 1822 AA.

AC R55273;
 DT 31-JAN-1995 (first entry)
 DE Beta subunit of integrin cell surface receptor.
 KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;
 KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;
 KW behaviour; signal transduction; receptor.

OS Homo sapiens.
 FH Key Location/Qualifiers

FT peptide 28..42
 FT modified_site 491 /note- "N-terminal peptide."

FT modified_site 617 /note- "Potential N-linked glycosylation site."
 FT modified_site 695 /note- "Potential N-linked glycosylation site."

FT domain /note- "Potential N-linked glycosylation site."
 FT 711..733
 FT /note- "Putative transmembrane domain."
 FT modified_site 980
 FT /note- "Potential N-linked glycosylation site."
 FT modified_site 1593 /note- "Potential N-linked glycosylation site."

PN US5320942-A.
 PD 14-JUN-1994.
 PF 19-FEB-1987; 016552.
 PR 19-FEB-1987; US-016552.
 PR 04-JAN-1989; US-293384.
 PR 01-OCT-1990; US-591105.

PA (KAT1/) KAT1 S.
 PA (QVAR/) QVARANTA V.
 PI Kaji S, Quaranta V;
 DR WPI: 94-191533/23.
 DR N-PSDB: 065674.

PT Diagnosing presence of abnormal epithelial tissue in vitro -
 PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein

PS Example 5; Figure 9; 34pp; English.
 CC Integrins are heterodimers comprised of alpha and beta subunits, that
 CC are non-covalently associated transmembrane glycoproteins. 11 alpha
 CC chains and 6 beta chains have been recognised in man. Each alpha
 CC subunit tends to associate with only one type of beta subunit but
 CC there are several exceptions. Integrins mediate (in part) the
 CC interaction of cells with the extracellular matrix, forming a link
 CC between the extracellular matrix and the cytoskeleton. They may
 CC transmit signals from the extracellular to the intracellular
 CC environment, affecting cell behaviour. This sequence is the beta4
 CC subunit of an alpha6 beta4 integrin.
 SQ Sequence 1822 AA;

Query Match 49.6%; Score 56; DB 1; Length 1822;
 Best Local Similarity 63.6%; Pred. No. 6.88e+01;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 650 ECFNKVKVND 660
 |||:|:|:|:|:|:|
 QY 2 ECFNKVKVND 12

RESULT 11
 ID R87107 standard; Peptide: 43 AA.

AC R87107;
 DT 28-AUG-1996 (first entry)
 DE Protocadherin clone RAT-224.

KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KW catenin; therapy; clone.
 OS Rattus rattus.

PN W03600289-A1.
 PD 04-JAN-1996.

PF 26-JUN-1995; U08071.
 PR 27-JUN-1994; US-268161.
 PA (DOHE-) DOHENY EYE INST.

PI Suzuki S;
 DR WPI: 96-068873/07.
 DR N-PSDB: T03582.

PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities

PS Example 1; Page 40; 146pp; English.

CC R87107-R87109 represent partial fragments of the rat protocadherin
 CC sequence. The cDNAs encoding these sequences were isolated after
 CC screening a rat brain cDNA preparation with the primers shown in T03575
 CC and T03576. The primers were constructed from portions of the amino acid
 CC sequences of the third and fourth extracellular domains of published
 CC cadherin sequences. The full length cDNA sequence encoding rat
 CC protocadherin pc5 is represented in T03574. The cytoplasmic domain of

CC cadherin interacts with the cytoskeleton through catenins and other
 CC cytoskeleton associated proteins. The cytoplasmic domain is not present
 CC in all cadherins, but in those which possess it, it is essential for the
 CC cadherin adhesive function. The cadherins which do not possess a
 CC cytoplasmic domain appear to function via a different method from those

CC with a cytoplasmic domain. These protein sequences are involved in
 CC cell-cell adhesion. These sequences may have regulatory functions in the
 CC cell, as well as the cell-cell adhesive properties. Antibodies produced
 CC against these sequences are useful for modulating the binding activity of
 CC these proteoglycans, and can be used therapeutically.
 SO Sequence 43 AA:

Query Match 47.8% Score 54: DB 1: Length 43;
 Best Local Similarity 66.7% Pred. No. 1.09e+02;
 Matches 10: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 29 CKIKVVDVNDNAP 43
 | ||| ||||| |
 Oy 1 CECNIKVDVNDNFP 15

RESULT 12
 ID R58865 standard; Protein: 43 AA.
 AC R58865:
 DT 17-APR-1995 (first entry)
 DE Rat-224 cadherin partial sequence.
 KW Cadherin: cell adhesion molecule.
 PN MO9414960-A.
 PD 07-JUL-1994.
 PR 23-DEC-1993: U12588.
 PR 29-DEC-1992: US-998003.
 PA (DOHE-) DOHEM EYE INST.
 PI Suzuki S;
 DR WPI: 94-293849/36.
 DR N-PSDB: 068957.
 PT Polynucleotide sequences encoding new proto-cadherins - useful
 PT for modulating natural binding and regulating activities.
 PS Example: Page 38: 114pp. English.
 CC Two regions of conserved AA sequence, one from the middle of the
 CC third cadherin extracellular subdomain (EC-3) and the other from the
 CC C-terminus of the fourth extracellular subdomain (EC-4) were
 CC identified. The corresp. degenerate oligos (Q68949, Q68950) were
 CC designed for use as PCR primers. PCR was carried out on a rat brain
 CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.
 CC The 450 bp band corresponded to the expected length between the two
 CC primer sites, but the 130 bp band could not be predicted from any
 CC of the previously identified cadherin sequences. The 450 bp and 130
 CC bp bands were extracted and sequenced. Nineteen novel partial cDNA
 CC clones were isolated. The DNA and deduced AA sequences of the
 CC clones (including sequences corresp. to the PCR primers) are given
 CC in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the
 CC cDNA clones are homologous to, but distinct from the known
 CC cadherins. The cadherins described thus far have highly conserved
 CC short AA sequences in the EC-3 including the consensus sequence
 CC D-Y-E or D-E-F located at the middle region of the subdomain and
 CC the consensus sequence in R58879 or R58880 at its end, while the
 CC corresp. sequences of other subdomains, except for the 5th extra-
 CC cellular subdomain (EC-5), are D-R-E and the sequence in R58881
 CC respectively. In contrast the deduced AA sequences of the new
 CC clones that corresp. to cadherin extracellular subdomains include
 CC the sequence D-Y-E or D-F-E at one end, but have the sequence
 CC D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.
 CC The polypeptides encoded by the partial clones are homologous to
 CC previously identified cadherins but did not show significant
 CC homology to any other sequences in Genbank. Therefore, the partial
 CC cDNAs appear to comprise a new subclass of cadherin-related
 CC molecules.
 SO Sequence 43 AA:

Query Match 47.8% Score 54: DB 1: Length 43;
 Best Local Similarity 66.7% Pred. No. 1.09e+02;
 Matches 10: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 29 CKIKVVDVNDNAP 43
 | ||| ||||| |
 Oy 1 CECNIKVDVNDNFP 15

RESULT 13
 ID W55470 standard; Protein: 83 AA.

AC W55470:
 DT 24-JUN-1998 (first entry)
 DE H. pylori ORF 06ap1119.16594193.fl.9 secreted protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW Identification; binding compound; bacteria; life cycle; activator;
 KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN MO9737044-A1.
 PD 09-OCT-1997.
 PR 27-MAR-1997: U05223.
 PR 06-DEC-1996: US-761218.
 PR 29-MAR-1996: US-625811.
 PR 02-APR-1996: US-758731.
 PR 25-OCT-1996: US-736905.
 PR 28-OCT-1996: US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB: V24879.

PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 94; Page 677-678; 1145pp; English.
 CC This sequence is a H. pylori secreted protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SO Sequence 83 AA:

Query Match 46.9% Score 53: DB 1: Length 83;
 Best Local Similarity 50.0% Pred. No. 1.37e+02;
 Matches 5: Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 43 CENCNIKIPSI 52
 | : ||||| :
 Oy 1 CECNIKVDV 10

RESULT 14
 ID W55280 standard; Protein: 87 AA.
 AC W55280:
 DT 02-JUL-1998 (first entry)
 DE H. pylori ORF 11ge10309orf7 protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW Identification; binding compound; bacteria; life cycle; activator;
 KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN MO9737044-A1.
 PD 09-OCT-1997.
 PR 27-MAR-1997: U05223.
 PR 06-DEC-1996: US-761318.
 PR 29-MAR-1996: US-625811.
 PR 02-APR-1996: US-758731.
 PR 25-OCT-1996: US-736905.

OY : : 1:|||||
5 IKVKDVNDNFP 15
Search completed: Sat May 13 07:22:09 2000
Job time : 7 secs.

PR 28-OCT-1996: US-738859.
PA (ASTR) ASTRA AB.
PI Alm RA, Smith D;
DR WPI: 97-503122/46.
DR N-PSDB: V24689.
PT Helicobacter pylori nucleic acid sequences and encoded polypeptide(s) - useful in vaccines to treat or prevent H. pylori infection and for diagnosis of H. pylori infection
PS Claim 14: Page 516; 1145pp; English.
CC This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds.
CC useful as potential H. pylori life cycle activators or inhibitors. The DNA and probes derived from it may be used for the identification of H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic acid sequences complementary to the DNA act as antisense sequences and can be used to prevent the translation of H. pylori mRNA. Antibodies against the protein can be used in immunoassays to evaluate the abundance and distribution of H. pylori-specific antigens. The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.
SQ Sequence 87 AA;

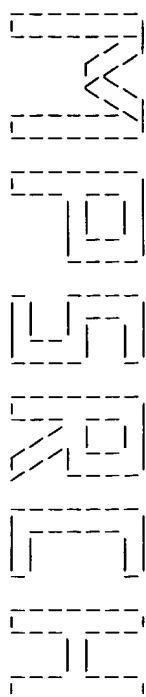
Query Match 46.9%; Score 53; DB 1; Length 87;
Best Local Similarity 50.0%; Pred. No. 1.37e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 47 CNCNIKIPST 56
1:|||||:
OY 1 CECNIKVKDV 10

RESULT 15
ID W13139 standard; Protein: 11 AA.
AC W13139;
DT 14-MAY-1997 (first entry)
DE Human cadherin-5 antigenic epitope (residues 242-252).
KW Ca2+ dependent; cell adhesion protein; cadherin; human; antibody;
KW purification; determination; epitope; tissue expression;
KW binding antagonist; calcium ion; antigen.
OS Homo sapiens.
PN US5597725-A.
PD 28-JAN-1997.
PE 17-APR-1992; 872643.
PR 17-APR-1992; US-872643.
PR 19-APR-1993; US-049460.
PR 26-JAN-1994; US-188228.
PA (DOHE-) DOHEMY EYE INST.
PI Suzuki S;
DR WPI: 97-108328/10.
PT Antibodies to cadherin proteins - useful as cadherin antagonists, etc.
PS Claim 5; Column 112; 59pp; English.
CC The present sequence is an antigenic epitope from human cadherin-5, which is a Ca2+ dependent cell adhesion protein. Antibodies or fragments that specifically bind the epitope can be used to purify the cadherin, determine its tissue expression and antagonise its ligand/antiligand binding activities.
SQ Sequence 11 AA;

Query Match 46.0%; Score 52; DB 1; Length 11;
Best Local Similarity 54.5%; Pred. No. 1.72e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 1 VTLDINDNFP 11



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:21:36 2000; MasPar time 4.30 Seconds
164.428 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNKKVADVNDNP 15

Scoring table: PAM 150
Gap 15

Searched: 142080 segs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 25.922; Variance 35.935; scale 0.721

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	113	100.0	999	1	IJHUG3	desmoglein 3 precursor 1.81e-12
2	89	78.8	1043	1	IJBOG1	desmoglein 1 precursor 9.39e-07
3	89	78.8	1049	1	IJHUG1	desmoglein 1 precursor 9.39e-07
4	66	58.4	562588	1	IJHUG1	desmoglein 1 precursor 9.39e-07
5	63	55.8	394	2	H64448	polyferredoxin - Meth 8.66e-02
6	60	53.1	809	1	IJBODD	desmocollin 2b precursor 1.40e-01
7	60	53.1	863	1	IJBODD	desmocollin 2a precursor 1.28e+00
8	60	53.1	866	2	I45858	desmocollin - bovine 1.28e+00
9	59	52.2	709	2	T16176	hypothetical protein 1.98e+00
10	59	52.2	2163	2	T15276	hypothetical protein 1.98e+00
11	58	51.3	5276	2	S62484	hypothetical protein 3.05e+00
12	58	51.3	832	2	S55396	LI-cadherin - human 3.05e+00
13	58	51.3	1324	2	S51622	cut3 protein - fission 3.05e+00
14	57	50.4	235	2	E70378	DNA replication prote 4.66e+00
15	57	50.4	656	2	T10664	serine/threonine-spec 4.66e+00
16	57	50.4	682	2	T06106	proliferating cell nu 4.66e+00
17	57	50.4	794	1	IJHUC5	cadherin 5 precursor 4.66e+00
18	57	50.4	829	2	I46536	KSP-cadherin - rabbit 4.66e+00
19	57	50.4	907	1	IJHUBD	desmocollin 3a precursor 4.66e+00
20	57	50.4	941	1	IJHUBA	desmocollin 3a precursor 4.66e+00
21	57	50.4	1979	2	C71622	hypothetical protein 4.66e+00
22	56	49.6	413	1	DEAGNT	D-nopaline dehydriogen 7.09e+00
23	56	49.6	787	2	S68699	potassium channel pro 7.09e+00

24	56	49.6	964	2	JC5545	integrin beta4E - hum 7.09e+00
25	56	49.6	1807	2	JC6319	integrin beta-4 chain 7.09e+00
26	56	49.6	1875	2	A36429	integrin beta-4 chain 7.09e+00
27	55	48.7	296	2	A71232	hypothetical protein 1.07e+01
28	55	48.7	376	2	E70361	chaperone DnaJ - Aqu 1.07e+01
29	55	48.7	496	2	H75122	sarcosine oxidase, ch 1.07e+01
30	55	48.7	902	2	T00588	hypothetical protein 1.07e+01
31	55	48.7	1389	2	I50090	carboxypeptidase gp18 1.07e+01
32	55	48.7	108	2	C72231	ATP synthase F1, subu 1.62e+01
33	54	47.8	249	2	G64415	hypothetical protein 1.62e+01
34	54	47.8	493	2	E71008	hypothetical protein 1.62e+01
35	54	47.8	671	2	A45730	phenol 2-monooxygenas 1.62e+01
36	54	47.8	807	2	T12177	potassium channel pro 1.62e+01
37	54	47.8	1790	2	S67593	transport protein USO 1.62e+01
38	53	46.9	265	2	G72340	conserved hypotheticala 2.42e+01
39	53	46.9	302	2	T02480	sec13-related protein 2.42e+01
40	53	46.9	380	2	S49116	hypothetical protein 2.42e+01
41	53	46.9	444	1	WHRTW	tryptophan 5-monooxyg 2.42e+01
42	53	46.9	646	2	B70396	histidine kinase sens 2.42e+01
43	53	46.9	836	2	A69550	hypothetical protein 2.42e+01
44	53	46.9	988	1	DJVEZP	DNA-directed DNA poly 2.42e+01
45	53	46.9	1986	2	S28353	probable polyketide s 2.42e+01

ALIGNMENTS

RESULT 1
ENTRY IJHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal name Homo sapiens #common name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
REFERENCE A41088
#authors Amagel, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references W01D:92069753
#accession A41088
#molecule_type mRNA
#residues 1-999 #label AMA
#cross-references GB:M76482; NID:g190751; PIDN:AAA60230.1; PID:g190754

GENETICS
#gene GDB:DSG3
#cross-references GDB:134030; OMIM:169615
#map_position 18q12.1-18q12.2
CLASSIFICATION
#superfamily cadherin: cadherin repeat homology
KEYWORDS
#protein binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE

1-23
24-49
50-999
50-615
52-157
160-267
270-333
390-455
496-588
616-639
640-999
910-938
937-966
110,180,545
#domain signal sequence #status predicted #label SIG\
#domain propeptide #status predicted #label PRO\
#product desmoglein homolog #status predicted #label MAT\
#domain extracellular #status predicted #label EXT\
#domain cadherin repeat homology #label CR1\
#domain cadherin repeat homology #label CR2\
#domain cadherin repeat homology #label CR3\
#domain cadherin repeat homology #label CR4\
#domain cadherin repeat homology #label CR5\
#domain transmembrane #status predicted #label TMN\
#domain intracellular #status predicted #label INT\
#domain desmoglein repeat #label DG1\
#domain desmoglein repeat #label DG2\
#binding site carbonylstate (Asn) (covalent) #status predicted
#length 999 #molecular-weight 107502 #checksum 8311

SUMMARY
Query Match 100.0%; Score 113; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.81e-12;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 251 CECNIKVDVNDNP 265
 OY 1 CECNIKVDVNDNP 15

RESULT 2

ENTRY 1JB0C1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #formal_name Bos primigenius laurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
 #authors S14603: A38872: A37785: S38721: A48173: S24412
 S14603
 #journal Koch, P.J.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.
 #submission Submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule-type mRNA
 #residues 1-1043 ##label KO2

REFERENCE
 #cross-references EMBL:X58466; NID:g306; PIDN:CAA41380.1; PID:g307
 #authors A38872
 Koch, P.J.: Goldschmidt, M.D.: Walsh, M.J.: Zimbelmann, R.: Franke, W.W.

#journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 #molecule-type mRNA
 #residues 1-87:968-1043 ##label KO2

REFERENCE
 #cross-references GB:S64268; GB:S64270
 #authors A37785
 Goodwin, L.: Hill, J.E.: Raynor, K.: Raszi, L.: Manabe, M.: Cowlin, P.

#journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 #molecule-type mRNA
 #residues 44-123, 'V', 125-493 ##label GOO

REFERENCE
 #cross-references GB:M58165; NID:g162966; PIDN:AAA62709.1; PID:g552318
 #authors S38721
 Zimbelmann, R.

#submission Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule-type mRNA
 #residues 44-1043 ##label ZIM

#cross-references EMBL:X57784; NID:g436061; PIDN:CAA0930.1; PID:g436062

REFERENCE
 #authors A48173
 Koch, P.J.: Walsh, M.J.: Schmelz, M.: Goldschmidt, M.D.: Zimbelmann, R.: Franke, W.W.

#journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references MUID:91168965
 #accession A48173
 #molecule-type mRNA
 #residues 44-1001, 'AOPPSAT' ##label KO3

#cross-references GB:X57784
 #accession A48173
 #note This sequence has been revised in references A38872 and S38721

GENETICS
 #gene DSG1
 #classification superfamily cadherin; cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein;

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110

transmembrane protein

#domain signal sequence #status predicted #label SIG
 #domain propeptide #status predicted #label PRO
 #product desmoglein #status predicted #label MAT
 #domain extracellular #status predicted #label EXT
 #domain cadherin repeat homology #label CR1
 #domain cadherin repeat homology #label CR2
 #domain cadherin repeat homology #label CR3
 #domain cadherin repeat homology #label CR4
 #domain transmembrane #status predicted #label TM1
 #domain intracellular #status predicted #label INT
 #domain desmoglein repeat #label DG1
 #domain desmoglein repeat #label DG2
 #domain desmoglein repeat #label DG3
 #domain desmoglein repeat #label DG4
 #region glycine/serine-rich
 #binding site carbohydrate (Asn) (covalent) #status experimental
 #binding site carbohydrate (Asn) (covalent) #status predicted
 #length 1043 #molecular-weight 112242 #checksum 6897

SUMMARY
 Query Match 78.8%; Score 89; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 9,39e-07;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 253 CECNIKVDVNDNP 267
 OY 1 CECNIKVDVNDNP 15

RESULT 3

ENTRY 1JHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

ACCESSIONS
 #authors S16906: A39706: A61254: A61279: S16158
 S16906
 #journal Buxton, R.S.

#submission Submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule-type mRNA
 #residues 1-1049 ##label BUX

REFERENCE
 #cross-references EMBL:X56554; NID:g30505; PIDN:CAA39976.1; PID:g30506
 #authors A39706
 Wheeler, G.N.: Parker, A.E.: Thomas, C.L.: Ataliotis, P.: Poynter, D.: Arnemann, J.: Rutman, A.J.: Pideley, S.C.: Watt, F.M.: Rees, D.A.: Buxton, R.S.: Magee, A.I.

#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 #molecule-type mRNA
 #residues 24-1049 ##label WHE

REFERENCE
 #cross-references GB:X56554
 #authors A61254
 Nilles, L.A.: Parry, D.A.D.: Powers, E.E.: Angst, B.D.: Wagner, R.M.: Green, K.J.

#journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#cross-references MUID:92121251
 #accession A61254
 #molecule-type mRNA
 #residues 26-1049 ##label NTL

REFERENCE
 #authors A61279
 Wheeler, G.N.: Buxton, R.S.: Parker, A.E.: Arnemann, J.: Rees, D.A.: King, I.A.: Magee, A.I.

```

#journal      Biochem. Soc. Trans. (1991) 19:1060-1064
#title        Desmosomal glycoproteins I, II and III: novel members of the
#             cadherin superfamily.
#cross-references WUID:92175187
#accession    A61279
#status       not compared with conceptual translation
#molecule_type mRNA
#residues     1-55 ##label WH3

GENETICS
#gene         GDB:DSG1
#cross-references GDB:126563; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION
#superfamily  #status predicted #label PRO\
#product       calcium binding; cell adhesion; duplication; glycoprotein;
#keywords      Transmembrane protein

FEATURE
1-23          #domain signal sequence #status predicted #label SIG\
24-49         #domain propeptide #status predicted #label PRO\
50-1049       #product desmoglein #status predicted #label M7\
50-548        #domain extracellular #status predicted #label EXT\
52-157        #domain cadherin repeat homology #label CR1\
160-269       #domain cadherin repeat homology #label CR2\
272-385       #domain cadherin repeat homology #label CR3\
393-493       #domain cadherin repeat homology #label CR4\
509-530       #region serine/threonine-rich\
549-569       #domain intracellular #status predicted #label TM\
572-1049      #domain desmoglein repeat #label DG1\
840-869       #domain desmoglein repeat #label DG2\
870-899       #domain desmoglein repeat #label DG3\
900-927       #domain desmoglein repeat #label DG4\
928-956       #region glycine/serine-rich\
969-1019      #binding-site carbohydrate (Asn) (covalent) #status
110,180       Predicted

SUMMARY
#length 1049 #molecular-weight 113715 #checksum 4482

Query Match      78.8% Score 89: DB 1: Length 1049:
Best Local Similarity 80.0% Pred. No. 9,396-07:
Matches 12: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

Db 253 CECNIKVDVNDNF 267
      |||||:|||||:
Oy 1 CECNIKVDVNDNF 15

RESULT 4
ENTRY      S62588 #type complete
TITLE      hypothetical protein SPAC1E11.03c - fission yeast
            (Schizosaccharomyces pombe)
ORGANISM   #formal_name Schizosaccharomyces pombe
            16-May-1996 #sequence_revision 13-Mar-1997 #text_change
DATE       31-Oct-1997

ACCESSIONS S62588
REFERENCE  S62586
#authors    McLean, J.; Harris, D.
#submission submitted to the EMBL Data Library, November 1995
#accession  S62588
#status     Preliminary
#molecule_type DNA
#residues   1-171 ##label MCL
#cross-references EMBL:Z67999; NID:91067216; PID:91067219

GENETICS
#map_position 1L
CLASSIFICATION
#superfamily  #status predicted #label FND
#product       domain fos/jun DNA-binding domain homology #label FND
#keywords      length 171 #molecular-weight 19348 #checksum 9337

SUMMARY
Query Match      58.4% Score 66: DB 2: Length 171:
Best Local Similarity 35.7% Pred. No. 8,666-02:
Matches 5: Conservative 6: Mismatches 3: Indels 0: Gaps 0:

Db 76 COCSVKINSVLTDF 89

```

```

Oy 1 CECNIKVDVNDNF 14
      |::|::|::|
RESULT 5
ENTRY      H64448 #type complete
TITLE      polyferredoxin - Methanococcus jannaschii
ORGANISM   #formal_name Methanococcus jannaschii
            13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
DATE       11-Jun-1999

ACCESSIONS H64448
REFERENCE  H64300
#authors    Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
            R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
            R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
            Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
            Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
            Scott, J.L.; Geoghegan, N.S.M.; Meldrum, J.F.; Fuhrmann,
            J.L.; Nguyen, D.; Uitterlinden, T.R.; Kelley, J.M.; Peterson,
            J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
            K.M.; Haurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
            H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
            Science (1996) 273:1058-1073
            Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii.

#journal     Methanococcus jannaschii.
#title       Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii.
#cross-references WUID:9637999
#accession   H64448
#status      preliminary: nucleic acid sequence not shown:
            translation not shown

#molecule_type DNA
#residues     1-394 ##label BUL
#cross-references GB:U677117; NID:91591813; PIDN:AA899195.1;
            #cross-references GB:U67860; TIGR:M01193; PID:91511192
            PID:91591821; TIGR:M01193; PID:91511192

GENETICS
#map_position FOR134591-113575
CLASSIFICATION
#superfamily  polyferredoxin 6x2(4Fe-4S); ferredoxin 2(4Fe-4S)
            homology

FEATURE
6-34        #domain ferredoxin 2(4Fe-4S) homology #label FER1\
66-119      #domain ferredoxin 2(4Fe-4S) homology #label FER2\
135-189     #domain ferredoxin 2(4Fe-4S) homology #label FER3\
203-258     #domain ferredoxin 2(4Fe-4S) homology #label FER4\
272-324     #domain ferredoxin 2(4Fe-4S) homology #label FER5\
336-389     #domain ferredoxin 2(4Fe-4S) homology #label FER6\

SUMMARY
#length 394 #molecular-weight 43342 #checksum 4356

Query Match      55.8% Score 63: DB 2: Length 394:
Best Local Similarity 53.8% Pred. No. 3,406-01:
Matches 7: Conservative 3: Mismatches 3: Indels 0: Gaps 0:

Db 83 CPGIRKVDVNDNF 95
      |::|::|::|
Oy 3 CNIKVDVNDNF 15

RESULT 6
ENTRY      I1B0DD #type fragment
TITLE      desmocollin 2b precursor - bovine (fragment)
            epithelial type 2 desmocollin subform II
ORGANISM   #formal_name Bos primigenius taurus #common_name cattle
            30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
DATE       22-Jun-1999

ACCESSIONS B41799
REFERENCE  A41799
#authors    Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky,
            R.; Franke, W.W.
            Proc. Natl. Acad. Sci. U.S.A. (1992) 89:353-357
            Complexity and expression patterns of the desmosomal
            cadherins.
#journal     cadherins.
#title       Complexity and expression patterns of the desmosomal
            cadherins.
#cross-references WUID:92108053
#accession   B41799
#molecule_type mRNA
#residues     1-809 ##label KOC

```


Best Local Similarity 54.5%: Pred. No. 1.98e+00;
Matches 6: Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 375 DIKIKVGESEF 385
QY 4 NIKVKVDVNDNF 14

RESULT 10
ENTRY T15276 #type complete
TITLE hypothetical protein R10F2.1 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS T15276
REFERENCE 218320

#authors Du, Z.; Gattung, S.
#submision submitted to the EMBL Data Library, May 1997
#description The sequence of C. elegans cosmid R10F2.
#accession T15276

#status preliminary; translated from GB/EMBL/DBJ

#molecule_type DNA

#residues 1-2163 #label DUZ

#cross-references EMBL:AF003388; NID:q2088850; PID:q2088852;
#experimental_source strain Bristol N2; clone R10F2

GENETICS
#gene CESP:R10F2.1

#map_position 3
#introns 34/1; 114/3; 735/3; 1552/2; 1800/3; 1865/3; 1911/3; 2026/3;
2108/2; 2150/3

SUMMARY #length 2163 #molecular-weight 238609 #checksum 7114

Query Match 52.2%: Score 59; DB 2; Length 2163;
Best Local Similarity 53.3%: Pred. No. 1.98e+00;

Matches 8: Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1629 CKCHVILVDNDLP 1643
QY 1 CECNIKVKVDVNDNF 15

RESULT 11
ENTRY S62484 #type complete
TITLE hypothetical protein SPAC4G8.07c - fission yeast
ORGANISM (Schizosaccharomyces pombe)
#formal_name Schizosaccharomyces pombe
DATE 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Oct-1997

ACCESSIONS S62484
REFERENCE S62445

#authors Badcock, K.; Churcher, C.M.
#submision submitted to the EMBL Data Library, October 1995
#accession S62484

#status preliminary

#molecule_type DNA
#residues 1-527 #label BAD
#cross-references EMBL:z56276; NID:g1022345; PID:g1022352

GENETICS

#map_position 1L
#introns 125/1; 158/2; 437/3; 520/3
SUMMARY #length 527 #molecular-weight 59613 #checksum 8195

Query Match 51.3%: Score 58; DB 2; Length 527;
Best Local Similarity 42.9%: Pred. No. 3.05e+00;

Matches 6: Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 275 ECPATKTINEEYP 288
QY 2 ECNIKVKVDVNDNF 15

RESULT 12

ENTRY S55396 #type complete
TITLE LI-cadherin - human
ORGANISM #formal_name Homo sapiens #common_name man

DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Sep-1997

ACCESSIONS S55396
REFERENCE S55396

#authors Boettlinger, A.; Kreft, B.; Flegler, C.; Dlouhy, B.; Berndorff, D.; Goessner, R.; Tauber, R.

#submision submitted to the EMBL Data Library, December 1994
#description Molecular cloning of human LI-cadherin: evidence for a novel type of cadherin within the cadherin superfamily.

#accession S55396

#status preliminary

#molecule_type mRNA
#residues 1-832 #label BOE
#cross-references EMBL:X83228; NID:g854174; PID:g854175

CLASSIFICATION #superfamily cadherin repeat homology
FEATURE 455-566

SUMMARY #domain cadherin repeat homology #label CR3
#length 832 #molecular-weight 92207 #checksum 9645

Query Match 51.3%: Score 58; DB 2; Length 832;
Best Local Similarity 64.3%: Pred. No. 3.05e+00;

Matches 9: Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 325 EHKVKVDVNDNF 338
QY 2 ECNIKVKVDVNDNF 15

RESULT 13
ENTRY S51622 #type complete
TITLE cut3 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 26-Aug-1999

ACCESSIONS S51622
REFERENCE S51622

#authors Saka, Y.; Sutani, T.; Yamashita, Y.; Satoh, S.; Takeuchi, M.; Nakaseko, Y.; Yanagida, M.

#journal EMBO J. (1994) 13:4938-4952
#title Fission yeast cut3 and cut4, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis.

#cross-references MUID:95045386

#accession S51622

#status preliminary

#molecule_type DNA
#residues 1-1324 #label SAK
#cross-references EMBL:D30788; NID:g577659; PID:d1007025; PID:g603501

CLASSIFICATION #superfamily conserved hypothetical P115 protein
SUMMARY #length 1324 #molecular-weight 150593 #checksum 3330

Query Match 51.3%: Score 58; DB 2; Length 1324;
Best Local Similarity 42.9%: Pred. No. 3.05e+00;

Matches 6: Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 200 CDVEITFEVNSDF 213
QY 1 CECNIKVKVDVNDNF 14

RESULT 14
ENTRY E70378 #type complete
TITLE DNA replication protein DnaC - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998

ACCESSIONS E70378
REFERENCE A70300

#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;

Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MIMD:98196666
#accession E70378
#status preliminary: nucleic acid sequence not shown;
translation not shown

##molecule-type DNA
##residues 1-235 #label AQF
##cross-references GB:AE000713; NID:g2983424; PID:g2983431; GB:AE000657
##experimental_source strain VFS

GENETICS

SUMMARY #gene dnaC #length 235 #molecular-weight 26934 #checksum 1332

Query Match 50.4%; Score 57; DB 2; Length 235;
Best Local Similarity 63.6%; Pred. No. 4.66e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 26 CECRFRKRDVN 36

OY 1 CECNIXVKDVN 11

RESULT 15

ENTRY T10664 #type complete
TITLE serine/threonine-specific protein kinase-like - Arabidopsis
thaliana

ALTERNATE_NAMES protein F6E21.20
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress

DATE 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change
20-Sep-1999

ACCESSIONS T10664
REFERENCE 216533

#authors

Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream,
M.A.; Barrell, B.G.; Bancroft, I.; Mewes, H.W.; Mayer,
K.F.X.; Lemcke, K.; Schellier, C.

#submission submitted to the Protein Sequence Database, June 1999

#accession T10664

##status preliminary

##molecule-type DNA

##residues 1-656 #label BEV

##cross-references EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.20

##experimental_source cultivar Columbia; BAC clone F6E21

GENETICS

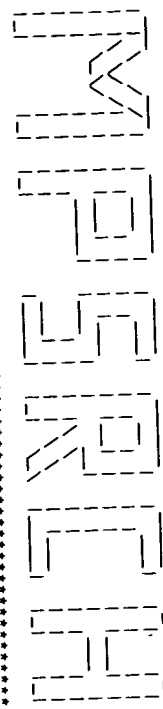
SUMMARY #gene ATSP:F6E21.20
#map_position 4
#introns 206/L: 243/L
#length 656 #molecular-weight 73503 #checksum 8670

Query Match 50.4%; Score 57; DB 2; Length 656;
Best Local Similarity 50.0%; Pred. No. 4.66e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 198 CECNLGYKDI 209

OY 1 CECNIXVKDVND 12

Search completed: Sat May 13 07:21:44 2000
Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:17:05 2000; Maspar time 66.90 Seconds
6.829 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table:
PAM 150
Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.676; Variance 31.455; scale 0.848

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	113	100.0	999	1	DSG3_HUMAN	DESMOGLIEN 3 PRECURSOR	5.22e-15
2	89	78.8	1043	1	DSG1_BOVIN	DESMOGLIEN 1 PRECURSOR	2.23e-08
3	89	78.8	1043	1	DSG1_HUMAN	DESMOGLIEN 1 PRECURSOR	2.23e-08
4	66	58.4	171	1	PCRI_SCHPO	TRANSSCRIPTION FACTOR P	1.19e-02
5	62	54.9	501	1	YONT_CAEEL	PUTATIVE G PROTEIN-COU	9.47e-02
6	60	53.1	863	1	DSC3_BOVIN	DESMOGLIEN 3A/2B PREC	2.60e-01
7	60	53.1	896	1	DSC3_BOVIN	DESMOGLIEN 3A/2B PREC	2.60e-01
8	59	52.2	709	1	SYR_CAEEL	PROBABLE ARGINYL-TRNA	4.27e-01
9	59	52.2	780	1	OBP_HSV6	REPLICATION ORIGIN BIN	4.27e-01
10	59	52.2	780	1	OBP_HSV6	REPLICATION ORIGIN BIN	4.27e-01
11	59	52.2	902	1	DSC2_MOUSE	DESMOGLIEN 2A/2B PREC	4.27e-01
12	58	51.3	527	1	VAD7_SCHPO	HYPOTHETICAL 59.6 KD P	6.96e-01
13	58	51.3	1324	1	CUT3_SCHPO	CHROMOSOME SEGREGATION	1.13e+00
14	57	50.4	783	1	CAD5_MOUSE	VASCULAR ENDOTHELIAL-C	1.13e+00
15	57	50.4	784	1	CAD5_MOUSE	VASCULAR ENDOTHELIAL-C	1.13e+00
16	57	50.4	901	1	DSC2_HUMAN	DESMOGLIEN 2A/2B PREC	1.82e+00
17	56	49.6	413	1	DHNO_AGR7	D-NOPALINE DEHYDROGENA	1.82e+00
18	56	49.6	782	1	CAD5_PIG	VASCULAR ENDOTHELIAL-C	1.82e+00
19	56	49.6	1807	1	ITB4_RAT	INTEGRIN BETA-4 PRECUR	1.82e+00
20	56	49.6	1875	1	ITB4_HUMAN	INTEGRIN BETA-4 PRECUR	1.82e+00
21	54	47.8	249	1	Y927_METJA	HYPOTHETICAL PROTEIN M	4.66e+00
22	54	47.8	670	1	TBUD_BURPI	PHENOL 2-MONOOXYGENASE	4.66e+00
23	54	47.8	1790	1	USOL_YEAST	INTRACELLULAR PROTEIN	4.66e+00

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	999 AA.
AC	DSG3_HUMAN			
AC	P32926			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	DESMOGLIEN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
CC	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
PA	Amagaki M., Klaus-Kovtun V., Stanley J.R.;			
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.";			
RT	Cell 67:869-877(1991).			
CC	- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.			
CC	- INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.			
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND CARCINOMAS.			
CC	- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).			
CC	- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES AGAINST DSG3.			
CC	- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED BY THE EMBL OUTSTATION - between the Swiss Institute of Bioinformatics and the EMBL OUTSTATION - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL: M76482; AAA60230.1; -			
DR	PIR: A41088; ITHUG3.			
DR	HSSP: P09803; 1EDH.			
DR	MM: 169615;			
DR	PROSITE: PS00232; CADHERIN: 3.			
DR	PFAM: PF00028; cadherin: 4.			
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;			

Calcium-binding; Repeat.
 1 23
 SIGNAL 1 23
 PROPEP 24 49
 CHAIN 50 999
 DOMAIN 50 615
 TRANSMEM 616 640
 DOMAIN 641 999
 REPEAT 50 158
 REPEAT 159 268
 REPEAT 269 383
 REPEAT 384 499
 REPEAT 499 935
 REPEAT 936 966
 CARBOHYD 110 110
 CARBOHYD 180 180
 CARBOHYD 459 459
 CARBOHYD 545 545
 SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;

Query Match 100.0%; Score 113; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 5,22e-15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

251 CECNIRKVDNDNFP 265
 1 CECNIRKVDNDNFP 15

RESULT 2
 ID DSG1-BOVIN STANDARD; PRT: 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 DS DGI.
 OS Bos taurus (Bovine).
 SC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 JC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 JC Bovinae; Bos.
 (1)
 RN SEQUENCE FROM N.A.
 RA TISSUE-MUZZLE EPITHELIUM;
 RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Franke W.W.;
 RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RX MEDLINE: 91168965.
 RA Koch P.J., Walsh M.J., Schmelz M., Goldschmidt M.D.,
 RA Zimbelmann R., Franke W.W.;
 RT Identification of desmoglein, a constitutive desmosomal
 RT glycoprotein, as a member of the cadherin family of cell adhesion
 RT molecules.
 RL Eur. J. Cell Biol. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RX MEDLINE: 92037656.
 RA Koch P.J., Goldschmidt M.D., Walsh M.J., Zimbelmann R.,
 RA Franke W.W.;
 RT Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.
 RL Eur. J. Cell Biol. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RX MEDLINE: 91097553.
 RA Goodwin L., Hill J.E., Raynor K., Raszi L., Manabe M., Cowin P.;
 RT Desmoglein shows extensive homology to the cadherin family of cell
 RT adhesion molecules.
 RA Blochem. Biophys. Res. Commun. 173:1224-1230(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X58466; CAA41380.1;
 CC EMBL: X57784; CAA40930.1;
 CC EMBL: M58165; AAA62709.1;
 CC PIR: S14603; IJBOG1.
 CC HSP: P09803; IEDH.
 CC PROSITE: P80023; CADHERIN; 2.
 CC PFM: PFM0028; cadherin; 3.
 CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
 CC Calcium-binding; Repeat.
 CC
 CC SIGNAL 1 23
 PROPEP 24 49
 CHAIN 50 1043
 DOMAIN 50 548
 TRANSMEM 549 573
 DOMAIN 574 1043
 REPEAT 159 270
 REPEAT 271 385
 REPEAT 386 498
 REPEAT 499 845
 REPEAT 846 875
 REPEAT 876 905
 REPEAT 906 933
 REPEAT 934 962
 REPEAT 963 1012
 CARBOHYD 110 110
 CARBOHYD 180 180
 CARBOHYD 496 496
 CONFLICT 124 124
 SEQUENCE 1043 AA: 112243 MW: ADE46133F8877C11 CRC64;
 Best Local Similarity 78.8%; Score 89; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 2.23e-08;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

253 CECNIRKVDNDNFP 267
 1 CECNIRKVDNDNFP 15

RESULT 3
 ID DSG1-HUMAN STANDARD; PRT: 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 DS DGI.
 OS Homo sapiens (Human).
 SC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 JC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 (1)
 RN SEQUENCE FROM N.A.
 RA TISSUE-KERATINOCYTES;
 RX MEDLINE: 91271279.
 RA Wheeler G.N., Parker A.E., Thomas C.L., Ataliotis P., Poynter D.,
 RA Arneimann J., Rutan A.V., Pidsley S.C., Watt F.M., Rees D.A.,
 RA Buxton R.S., Magee A.I.;
 RT Desmosomal glycoprotein DGI, a component of intercellular desmosome

RT junctions, is related to the cadherin family of cell adhesion molecules.":

RT Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X56554; CAA39976.1; -

DR PIR: S16906; IJHUG1.

DR HSSP: P09803; IEDH.

DR MIM: 125670; -

DR PROSITE: PS00332; CADHERIN: 2.

DR PFM: PF00028; cadherin: 4.

KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; Calcium-binding; Repeat.

KM SIGNAL 1 23

FT PROPEP 24 49 POTENTIAL.

FT CHAIN 50 1049 DESMOGLEIN 1.

FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 546 570 POTENTIAL.

FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).

FT REPEAT 50 158 CADHERIN 1.

FT REPEAT 159 270 CADHERIN 2.

FT REPEAT 271 385 CADHERIN 3.

FT REPEAT 386 497 CADHERIN 4.

FT REPEAT 813 839 DESMOGLEIN REPEAT 1.

FT REPEAT 840 869 DESMOGLEIN REPEAT 2.

FT REPEAT 870 899 DESMOGLEIN REPEAT 3.

FT REPEAT 900 927 DESMOGLEIN REPEAT 4.

FT REPEAT 928 956 DESMOGLEIN REPEAT 5.

FT DOMAIN 969 1019 GLY/SER-RICH.

FT CARBOHYD 36 36 POTENTIAL.

FT CARBOHYD 110 110 POTENTIAL.

FT CARBOHYD 180 180 POTENTIAL.

SO SEQUENCE 1049 AA: 113715 MW: 12565589D6619 CRC64:

Query Match 78.8%; Score 89; DB 1; Length 1049;

Best Local Similarity 80.0%; Pred. No. 2,23e-08;

Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 253 CECNIIKLVNDNIP 267

OY 1 CECNIIKLVNDNIP 15

RESULT 4

ID PCRI-SCHPO STANDARD: PRT: 171 AA.

AC 009926;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE TRANSCRIPTION FACTOR PCRI (TRANSCRIPTION FACTOR MTS2).

GN PCRI OR MTS2 OR SPAC21E11.03C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

CC Schizosaccharomycetaceae; Schizosaccharomycetes.

CC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE: 9610438.

RX Matsuoka Y., Yamamoto M.;

RT "Schizosaccharomyces pombe pcrl+ encodes a CREB/ATF protein involved in regulation of gene expression for sexual development.":

RT Mol. Cell. Biol. 16:704-711(1996).

CC [2]

RP SEQUENCE FROM N.A.

RA Kon N., Krawchuk M.D., Warren B.G., Smith G.R., Wahls W.P.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-972;

RL McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;

RN Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

RP CHARACTERIZATION.

RA MEDLINE: 95047325.

RX Wahls W.P., Smith G.R.;

RT "A heteromeric protein that binds to a meiotic homologous recombination hot spot: correlation of binding and hot spot activity."

RT Genes Dev. 8:1693-1702(1994).

CC -1- FUNCTION: INVOLVED IN REGULATION OF GENE EXPRESSION FOR SEXUAL DEVELOPMENT. BINDS AND ACTIVATES MEIOTIC RECOMBINATION HOT SPOT ADE6-M26.

CC -1- SUBUNIT: HETERODIMER OF PCRI/MTS2 AND ATF1/MTS1.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D63667; BAA0918.1; -

DR EMBL: U87870; ABA46981.1; -

DR EMBL: Z67999; CAA91968.1; -

DR HSSP: P05412; IFOS.

DR TRANSFAC: T01687; -

DR PROSITE: PS00036; BZIP_BASIC: 1.

DR PFM: PF00170; bzip: 1.

KW Transcription regulation; Activator; DNA-binding; Nuclear protein; Melosis.

FT DNA_BIND 12 32 BASIC MOTIF.

FT DOMAIN 42 66 LEUCINE-ZIPPER.

SO SEQUENCE 171 AA: 19348 MW: 99222FDDEE150BDE CRC64:

Query Match 58.4%; Score 66; DB 1; Length 171;

Best Local Similarity 35.7%; Pred. No. 1.19e-02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

DB 76 CCGSVKIRSVLTDF 89

OY 1 CCGSVKIRSVLTDF 14

RESULT 5

ID YONJ-CAEEL STANDARD: PRT: 501 AA.

AC 002213;

DT 15-FEB-2000 (Rel. 39, Created)

DT 15-FEB-2000 (Rel. 39, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE PUTATIVE G PROTEIN-COUPLED RECEPTOR C02D4.2.

GN C02D4.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

CC Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.

CC [1]

RN SEQUENCE FROM N.A.

RP STRAIN-BRISTOL N2;

RA Baynes C.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

```

CC -1- FUNCTION: PUTATIVE G-PROTEIN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC MOST SIMILAR TO INSECT OCTOPAMINE RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 281031; CAB02718.1; -
DR MORPEP: C0204.2; CE07837.
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR: 1.
DR PIR: PFO0001; 7tm_1; 2.
KM G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 49 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 50 72 1 (POTENTIAL).
FT TRANSSEM 73 82 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 83 104 2 (POTENTIAL).
FT DOMAIN 120 136 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 137 157 3 (POTENTIAL).
FT DOMAIN 158 177 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 178 200 4 (POTENTIAL).
FT DOMAIN 201 224 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 225 246 5 (POTENTIAL).
FT DOMAIN 247 373 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 374 395 6 (POTENTIAL).
FT DOMAIN 396 426 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 427 447 7 (POTENTIAL).
FT DOMAIN 448 487 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 493 496 POLY-ARG.
FT CARBOHYD 15 15 POTENTIAL.
FT DISULFID 134 212 BY SIMILARITY.
SQ SEQUENCE 501 AA: 56404 MW: 518140.6522888 CRC64:

Query Match 54.9%: Score 62; DB 1; Length 501;
Best Local Similarity 54.5%: Pred. No. 9.47e-02;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 336 CAYKVKDVKED 346
   1:::|||||:
OY 3 CNIKVKVDNDN 13

RESULT 6
ID DSC2_BOVIN STANDARD: PRT: 863 AA.
AC P33545;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN)
DE (FRAGMENT).
GN DSC2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE EPITHELIUM;
RA MEDLINE: 92108053.
RA Koch P.J., Goldschmidt M.D., Zimbelmann R., Troyanovsky R.,
RA Franke W.W.;
RA "Complexity and expression patterns of the desmosomal cadherins.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:353-357(1992).
RL -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.

```

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: FORMS 2A (SHOWN HERE) AND 2B ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: ESOPHAGUS AND RUMEN. WEAKLY IN EPITHELIA AND
CC CARDIAC MUSCLE.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -----
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81190; AAA30783.1; -
DR EMBL: M81190; AAA30783.1; -
DR PIR: A41799; IUBODC.
DR PIR: B41799; IUBODC.
DR HSSP: P09803; ISUH.
DR PROSITE: PS00232; CADHERIN: 3.
DR PIR: PFO0028; cadherin; 5.
KM Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW Alternative splicing; Cytoskeleton; Calcium binding.
FT NON_TER 1 1
FT PROPEP 1 89 POTENTIAL.
FT CHAIN 90 863 DESMOCOLLIN 2A/2B.
FT DOMAIN 90 644 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 645 665 POTENTIAL.
FT TRANSSEM 666 863 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 90 197 CADHERIN 1.
FT REPEAT 198 309 CADHERIN 1.
FT REPEAT 310 423 CADHERIN 2.
FT REPEAT 424 528 CADHERIN 3.
FT REPEAT 529 644 CADHERIN 4.
FT CARBOHYD 120 120 CADHERIN 5.
FT CARBOHYD 346 346 POTENTIAL.
FT CARBOHYD 495 495 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT VARIANT 264 264 K -> Q.
FT VARIANT 333 333 R -> D.
FT VARSPLIC 799 809 KVGQCDDQDQTH -> ESIHGHLVKN (IN ISOFORM
FT VARSPLIC 810 863 MISSING (IN ISOFORM 2B).
SQ SEQUENCE 863 AA: 95874 MW: 603854.0CA16727F2 CRC64:

Query Match 53.1%: Score 60; DB 1; Length 863;
Best Local Similarity 61.5%: Pred. No. 2.60e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 295 CIINIEDVNDNLP 307
   1:::|||||:
OY 3 CNIKVKVDVNDNFP 15

RESULT 7
ID DSC3_BOVIN STANDARD: PRT: 896 AA.
AC Q28060; Q28061; Q28176;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DESMOCOLLIN 3A/3B PRECURSOR.
GN DSC3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE EPITHELIUM;
RA MEDLINE: 95403557.
RA Yue K.K.M., Holton J.L., Clarke J.P., Hyam J.L.M., Hashimoto T.,

```

RA Chidgey M.A.J., Garrod D.R.:
 RT "Characterisation of a desmocollin isoform (bovine DSC3) exclusively
 RT expressed in lower layers of stratified epithelia.";
 RL J. Cell Sci. 108:2163-2173(1995).
 RN [2]
 RP SEQUENCE OF 686-814 FROM N.A.
 RC TISSUE-EPIDERMIS;
 RX MEDLINE: 94308280.
 RA Legan P.K., Yue K.K.M., Chidgey M.A.J., Holton J.L., Wilkinson R.W.,
 RA Garrod D.R.:
 RT "The bovine desmocollin family: a new gene and expression patterns
 RT reflecting epithelial cell proliferation and differentiation.";
 RL J. Cell Biol. 126:507-518(1994).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: FORMS 3A AND 3B ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE SAME GENE.
 CC -1- TISSUE SPECIFICITY: STRATIFIED EPITHELIA ONLY (EPIDERMIS, TONGUE,
 CC ESOPHAGUS AND RUMEN).
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L33774; AAC41625.1; -
 CC EMBL: L33774; AAC41626.1; -
 CC EMBL: X75783; CAA53427.1; -
 CC HSSP: P09803; ISUH.
 DR PRINTS: PR00205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PFAM: PF00028; cadherin. 5.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Alternative splicing; Cytoskeleton; Calcium-binding.
 FT SIGNAL 1 26
 FT PROPEP 27 134
 FT CHAIN 135 886
 FT DOMAIN 135 690
 FT TRANSMEM 691 711
 FT DOMAIN 712 896
 FT REPEAT 135 242
 FT REPEAT 243 354
 FT REPEAT 355 471
 FT REPEAT 472 579
 FT REPEAT 580 690
 FT CARBOHYD 165 165
 FT CARBOHYD 391 391
 FT CARBOHYD 546 546
 FT CARBOHYD 629 629
 FT VARSPLIC 832 839
 FT VARSPLIC 840 896
 FT CONFLICT 686 687
 SO SEQUENCE 896 AA: 99687 MW: 88CC0C30A63FB0BD4 CRC64:

Query Match 53.1% Score 60: DB 1: Length 896:
 Best Local Similarity 69.2% Pred. No. 2.60e-01;
 Matches 9: Conservative 1: Mismatches 3: Indels 0: Gaps 0;

DB 340 CITTAKDNDNLP 352
 1 1 1 1 1 1 1 1 1
 QY 3 CNIKKVDVNDNF 15

RESULT 8
 ID SVR_CAEEL STANDARD: PRT: 709 AA.
 AC Q19825;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE)
 DE (ARGRS).
 GN F26f4.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematozoa; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditidae; Rhabditidae; Pelodermidae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Fulton L.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARC) -> AMP +
 CC PYROPHOSPHATE + L-ARGINYL-TRNA(ARC).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U12964; AAA91224.1; -
 CC MORPEP: P26f4.10; CE01258.
 DR PRINTS: PR01036; TRNASYNTHARG.
 DR PROSITE: PS00178; AA-TRNA_LIGASE_1; 1.
 DR PFAM: PF00750; tRNA-synt_1d; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
 FT SIMILAR 252 263
 SO SEQUENCE 709 AA: 80477 MW: 175B7CAA6609689D CRC64:

Query Match 52.2% Score 59: DB 1: Length 709;
 Best Local Similarity 54.5% Pred. No. 4.27e-01;
 Matches 6: Conservative 4: Mismatches 1: Indels 0: Gaps 0;

DB 375 DIKIKVGESEF 385
 1 1 1 1 1 1 1 1 1
 QY 4 NIKKVDVNDNF 14

RESULT 9
 ID OBP_HSVGU STANDARD: PRT: 780 AA.
 AC P52378;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE REPLICATION ORIGIN BINDING PROTEIN (OBP).
 GN U73 OR 19R OR HDRPO.
 OS Herpes simplex virus (type 6 / strain Uganda-1102).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 95266321.
 RA Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
 RA Martin M.E., Efsthaliou S., Craxton M., Macculay H.A.;
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,
 RT and genome evolution.";
 RL Virology 209:29-51(1995).
 RN [2]
 RP SEQUENCE OF 1-703 FROM N.A.
 RC MEDLINE: 95146942.
 RA Lawrence G.L., Nicholas J., Barrell B.G.;
 RT "Human herpesvirus 6 (strain U1102) encodes homologues of the
 RT conserved herpesvirus glycoprotein gp and the alphaherpesvirus

```

RT origin-binding protein."
RL J. Gen. Virol. 76:147-152(1995).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC EHV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83413; CA58365.1; -.
DR EMBL: M68963; AAA65581.1; -.
KM DNA replication; DNA-binding; ATP-binding.
FT NP_BIND 52
SQ SEQUENCE 780 AA: 89716 MW: 53F4B0E4C9BCB5F CRC64:
Query Match 52.2%; Score 59; DB 1; Length 780;
Best Local Similarity 53.3%; Pred. No. 4,27e-01;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 424 CENSIKYDIVNGFP 438
OY 1 CENSIKYDIVNGFP 15
RESULT 10
ID OBP_HSV62 STANDARD: PRT: 780 AA.
AC P52452.1
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE REPLICATION ORIGIN BINDING PROTEIN (OBP).
OS U73 OR CHER.
GN Herpes simplex virus (type 6 / strain 229).
OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 96195263.
RA Lindquist G.J., Inoue N., Allen R.D., Castelli J.W.,
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,
RA Fenkel N., Pellett P.E.;
RT "Restriction endonuclease mapping and molecular cloning of the human
RT herpesvirus 6 variant B strain 229 genome."
RL Arch. Virol. 141:367-379(1996).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC EHV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L14772; AAB06356.1; -.
KM DNA replication; DNA-binding; ATP-binding.
FT NP_BIND 52
SQ SEQUENCE 780 AA: 89553 MW: 1DE68504FA9624B6 CRC64:
Query Match 52.2%; Score 59; DB 1; Length 780;
Best Local Similarity 53.3%; Pred. No. 4,27e-01;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 424 CENSIKYDIVNGFP 438

```

```

OY 1 CENSIKYDIVNGFP 15
RESULT 11
ID DSC2_MOUSE STANDARD: PRT: 902 AA.
AC P55292; 064734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN).
GN DSC2 OR DSC3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-C57BL/6; TISSUE-EMBRYO;
RX MEDLINE: 95227276.
RA Lorimer J.E., Hall L.S., Clarke J.P., Collins J.E., Fleming T.P.,
RA Garrod D.R.;
RT "Cloning, sequence analysis and expression pattern of mouse
RT desmocollin 2 (DSC2), a cadherin-like adhesion molecule."
RL Mol. Membr. Biol. 11:229-236(1994).
RN [2]
RP SEQUENCE OF 344-637 FROM N.A.
RC STRAIN-C57BL/6; TISSUE-EMBRYO;
RX MEDLINE: 95048328.
RA Buxton R.S., Wheeler G.N., Pidsley S.C., Marsden M.D., Adams M.J.,
RA Jenkins N.A., Gilbert D.J., Copeland N.G.;
RT "Mouse desmocollin (Dsc3) and desmoglein (Dsg1) genes are closely
RT linked in the proximal region of chromosome 18."
RL Genomics 21:510-516(1994).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: FORMS 2A AND 2B ARE PRODUCED BY ALTERNATIVE
CC SPLICING OF THE SAME GENE.
CC -1- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33779; AAA79177.1; -.
DR EMBL: L33779; AAA79176.1; -.
DR EMBL: X73885; CA52089.1; -.
DR HSP: P09803; IEDH.
DR MGD: MGI:103221; DSC2.
DR PRINTS: PR00205; CADHERIN.
DR PROSITE: PS00232; CADHERIN.
DR PFAM: PF00028; cadherin. 5.
KM Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KM Alternative splicing; Cytoskeleton; Calcium-binding.
FT SIGNAL 1
FT PROPEP 28 135
FT CHAIN 136 902
FT DOMAIN 136 694
FT TRANSMEM 695 715
FT DOMAIN 716 902
FT REPEAT 136 243
FT REPEAT 244 355
FT REPEAT 356 471
CADHERIN 3.

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X83930; CAA58782.1; -
 DR HSP: P09803; LEDH.
 DR MGD: MGI:105057; CDH5.
 DR PRINTS: PRO0205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PFAM: PF00028; cadherin. 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KW Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 24
 FT PROPEP 25 45 POTENTIAL.
 FT CHAIN 46 783 VASCULAR ENDOTHELIAL-CADHERIN.
 FT DOMAIN 46 592 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 593 619 POTENTIAL.
 FT DOMAIN 620 783 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 46 148 CADHERIN 1.
 FT REPEAT 149 255 CADHERIN 2.
 FT REPEAT 256 370 CADHERIN 3.
 FT REPEAT 371 475 CADHERIN 4.
 FT REPEAT 476 592 CADHERIN 5.
 FT DOMAIN 737 752 SER-RICH.
 FT CARBOHYD 59 59 POTENTIAL.
 FT CARBOHYD 154 154 POTENTIAL.
 FT CARBOHYD 440 440 POTENTIAL.
 FT CARBOHYD 522 522 POTENTIAL.
 FT CARBOHYD 534 534 POTENTIAL.
 SO SEQUENCE 783 AA: 87847 MW: 240A2D663BCE71C CRC64:
 Query Match 50.4%; Score 57; DB 1; Length 783;
 Best Local Similarity 63.6%; Pred. No. 1.13e+00;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 243 IRLDINDNFP 253
 I: I I I I I I I I
 Oy 5 IKVRVDNDFP 15

RESULT 15
 ID CAD5_HUMAN STANDARD: PRT: 784 AA.
 AC P33151;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
 DE (7B4 ANTIGEN) (CD144 ANTIGEN).
 GN CDH5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOTHELIAL CELLS:
 RX MEDLINE: 95353875.
 RA Breviario F., Cavada L., Corada M., Martin-Padura I., Navarro P.,
 RA Golay J., Introna M., Guino D., Lampugnani M.G., Dejana E.;
 RT "Functional properties of human vascular endothelial cadherin
 RT (7B4/cadherin-5), an endothelial-specific cadherin";
 RL Arterioscler. Thromb. Vasc. Biol. 15:1229-1239(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA:
 RX MEDLINE: 97362755.
 RA Ali J., Liao F., Martens E., Muller W.A.;
 RT "Vascular endothelial cadherin (VE-cadherin): cloning and role in
 RT endothelial cell-cell adhesion";
 RL Microcirculation 4:267-277(1997).

RN [3]
 RP SEQUENCE OF 5-784 FROM N.A.
 RC TISSUE-BRAIN:
 RX MEDLINE: 91283540.
 RA Suzuki S., Sano K., Tanihara H.;
 RT "Diversity of the cadherin family: evidence for eight new cadherins
 RT in nervous tissue.";
 RL Cell Regul. 2:261-270(1991).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC TISSUE-ENDOTHELIAL CELLS:
 RX MEDLINE: 92394977.
 RA Lampugnani M.G., Resnati M., Rauterl M., Pigott R., Pisecane A.,
 RA Houen G., Roco L.P., Dejana E.;
 RT "A novel endothelial-specific membrane protein is a marker of
 RT cell-cell contacts.";
 RL J. Cell Biol. 118:1511-1522(1992).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
 CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
 CC ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
 CC BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES.
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL TISSUES AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -1- DATABASE: NAME-PRO: NOTE-CD guide CD144 entry.
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd144.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79981; CAA56306.1; -
 DR EMBL: U84722; AAB41796.1; -
 DR EMBL: X59796; CAA42468.1; -
 DR PIR: S24305; ITHUC5.
 DR HSP: P09803; LEDH.
 DR MIM: 601120; -
 DR PRINTS: PRO0205; CADHERIN.
 DR PROSITE: PS00232; CADHERIN. 3.
 DR PFAM: PF00028; cadherin. 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
 KW Calcium-binding; Repeat; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 47 POTENTIAL.
 FT CHAIN 48 784 VASCULAR ENDOTHELIAL-CADHERIN.
 FT DOMAIN 48 593 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 594 620 POTENTIAL.
 FT DOMAIN 621 784 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 48 151 CADHERIN 1.
 FT REPEAT 152 258 CADHERIN 2.
 FT REPEAT 259 372 CADHERIN 3.
 FT REPEAT 373 477 CADHERIN 4.
 FT REPEAT 478 593 CADHERIN 5.
 FT DOMAIN 736 753 SER-RICH.
 FT CARBOHYD 61 61 POTENTIAL.
 FT CARBOHYD 112 112 POTENTIAL.
 FT CARBOHYD 157 157 POTENTIAL.
 FT CARBOHYD 362 362 POTENTIAL.
 FT CARBOHYD 442 442 POTENTIAL.
 FT CARBOHYD 523 523 POTENTIAL.
 FT CARBOHYD 535 535 POTENTIAL.
 FT CONFLICT 517 517 I -> T (IN REF. 1).
 SO SEQUENCE 784 AA: 87528 MW: 6DBEBFC4DA6899D1 CRC64:

Query Match 50.4% Score 57; DB 1; Length 784;
Best Local Similarity 81.8% Pred. No. 1.13e+00;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 139 IKVHDVNDNMP 149
||| ||||| |
QY 5 IKVKDVNDNFP 15

Search completed: Sat May 13 07:18:23 2000
Job time : 78 secs.

MIPED

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:18:42 2000: Maspar time 139.87 seconds

Tabular output not generated. 7.436 Million cell updates/sec

Title: >US-08-991-628-5

Description: (1-15) from US08991628.pep

Perfect Score: 113

Sequence: 1 CECNIKVKVDNDNFP 15

Scoring table: PAM 150

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08

Listing first 45 summaries

Database: sptemb12

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human

5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle

9:sp_phage 10:sp_plant 11:sp_protent 12:sp_unclassified

13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.749; Variance 32.815; scale 0.785

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	107	94.7	993	11	035902	DESMOGLEIN 3 (FRAGMENT	4.01e-12
2	72	63.7	787	4	09Y5E8	PROTODADHERIN BETA 15.	1.95e-03
3	65	57.5	222	3	013737	HYPOTHETICAL 24.9 KD P	7.00e-02
4	65	57.5	519	13	090425	VENTRAL NEURAL CADHERIN	7.00e-02
5	63	55.8	394	1	058593	METHYLOTOGEN-REDUCIN	1.88e-01
6	63	55.8	2606	5	021920	KLIA8.7 PROTEIN.	3.06e-01
7	62	54.9	379	2	008356	DNAJ PROTEIN.	3.06e-01
8	62	54.9	501	5	002213	PUTATIVE G PROTEIN-CON	3.06e-01
9	60	53.1	466	5	076322	SYNAPSIN S-SYN-SHORT (7.98e-01
10	60	53.1	503	5	076323	SYNAPSIN S-SYN-LONG (F	7.98e-01
11	60	53.1	814	6	077704	DESMOGLEIN TYPE 2 (FR	7.98e-01
12	59	52.2	265	5	09XU08	REPPLICATION ORIGIN-BIN	1.28e+00
13	59	52.2	659	14	069468	ORIGIN BINDING PROTEIN	1.28e+00
14	59	52.2	780	14	09WT02	SIMILARITY TO MULTIPLE	2.04e+00
15	59	52.2	2163	5	001912	MHC CLASS I PRECURSOR	2.04e+00
16	58	51.3	339	7	P79570	INTESTINAL PEPTIDE-ASS	2.04e+00
17	58	51.3	832	4	012864	LI-CADHERIN.	2.04e+00
18	58	51.3	832	4	015336	PEPTIDE SYNTHETASE.	2.04e+00
19	58	51.3	5137	2	001135	DNA REPLICATION PROTEI	3.24e+00
20	57	50.4	235	2	067056		

RESULT	ID	ALIGNMENTS	PRT	993 AA.
AC	035902	PRELIMINARY:	PRT:	993 AA.
DT	01-JAN-1998 (TREMblrel. 05, Created)			
DT	01-JAN-1998 (TREMblrel. 05, Last sequence update)			
DT	01-NOV-1999 (TREMblrel. 12, Last annotation update)			
DE	DESMOGLEIN 3 (FRAGMENT).			
GN	DSG3			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
CC	Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	ISHIKAWA H., LI K., Uitto J.			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).			
DR	EMBL: U86016; AAB65091.1; -			
DR	HSSP: P15116; INCI.			
DR	PROSITE: PS00232; CADHERIN; 2.			
DR	PFAM: PF00028; cadherin; 4.			
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.			
FT	NON_TER 993 993			
SO	SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;			

Query Match	Score 107; DB 11; Length 993;
Best Local Similarity 86.7%; Pred. No. 4.01e-12;	
Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	
Db 251 CECNIKVKVDNDNFP 265	
QY 1 CECNIKVKVDNDNFP 15	
RESULT 2	
ID 09Y5E8	PRELIMINARY; PRT; 787 AA.
AC 09Y5E8	
DT 01-NOV-1999 (TREMblrel. 12, Created)	
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)	
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)	
DE PROTODADHERIN BETA 15.	
GN PCDH-BETA15.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	

```

OC Eutheria: Primates: Catarrhini; Homiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 99308636.
RA WU O., MANIATIS T.:
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes."
RL Cell 97:779-790(1999).
RC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF152494; AAD43755.1; -.
DR PROSITE: PS00232; CADHERIN; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ SEQUENCE 787 AA; 86329 MW; F66574C0 CRC32;

Query Match 63.7% Score 72; DB 4; Length 787;
Best Local Similarity 66.7% Pred. No. 1.95e-03;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 331 C5SVKVLVDVNDNFP 345
OY 1 CECNKKVNDNFP 15

RESULT 3 PRELIMINARY: PRT; 222 AA.
ID 013737
AC 013737:
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 24.9 KD PROTEIN C16E8.02 IN CHROMOSOME 1.
GN SPAC16E8.02.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO YEAST YGL010W AND SOME, TO N.CRASSA ATP-6.
DR EMBL: Z85529; CAB1031.1; -.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT DOMAIN 30 33 POLY-LEU.
SQ SEQUENCE 222 AA; 24993 MW; 396C437E CRC32;

Query Match 57.5% Score 65; DB 3; Length 222;
Best Local Similarity 66.7% Pred. No. 7.00e-02;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 167 NIKLKDVNDNFP 178
OY 4 NIKVKVDNNDNFP 15

RESULT 4 PRELIMINARY: PRT; 519 AA.
ID 090425
AC 090425:
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE VENTRAL NEURAL CADHERIN (FRAGMENT).
GN VNC.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinoidae; Cyprinidae; Rasbortinae; Danio.
RN (1)

```

```

RP SEQUENCE FROM N.A.
RA FRANKLIN J.L., SARGENT T.D.;
RC Dev. Dyn. 206:0-0(0).
RL -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: U41419; AAC47406.1; -.
DR HSSP: P15116; 1NCJ.
DR ZFIN: ZDB-GENE-980526-170; vnc.
DR PROSITE: PS00232; CADHERIN; 1.
DR PFAM: PF00028; cadherin; 3.
DR PFAM: PF01049; cadherin_C-term; 1.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
SQ SEQUENCE 519 AA; 57807 MW; E6DA0079 CRC32;

Query Match 57.5% Score 65; DB 13; Length 519;
Best Local Similarity 91.7% Pred. No. 7.00e-02;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 202 NIKVKVDNNDNFP 213
OY 4 NIKVKVDNNDNFP 15

RESULT 5 PRELIMINARY: PRT; 394 AA.
ID 058593
AC 058593:
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE METHYLVIOLAGEN-REDUCING HYDROGENASE POLYFERREDOXIN PROTEIN.
GN VHUB OR M1193.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 2661;
RX MEDLINE: 96337899.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOGAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL Science 273:1058-1073(1996).
CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -1- CORFACTOR: BINDS TEN 4FE-4S CLUSTER.
CC -1- SUBUNIT: VHUB CONSISTS OF THE VHUA, VHUG, VHUU SUBUNTS AND A
CC FERREDOXIN PROTEIN.
CC -1- MISCELLANEOUS: THIS PROTEIN CONTAINS SEVERAL REPEATED FERREDOXIN-
CC LIKE DOMAINS.
DR EMBL: U67560; AAB99195.1; -.
DR HSSP: P00185; 1CLF.
DR TIGR: M01193; -.
DR PFAM: PF00037; fer4; 6.
DR PFAM: PF00353; 4FE4SFRDOXIN.
KW Electron transport; Iron-sulfur; Repeat.
FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 26 26 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 33 33 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 36 36 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 39 39 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 42 42 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 46 46 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 49 49 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 52 52 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 56 56 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).

```

```

FT METAL 79 79 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)
FT METAL 83 83 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY)
FT METAL 101 101 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)
FT METAL 104 104 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)
FT METAL 107 107 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)
FT METAL 111 111 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY)
FT METAL 142 142 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY)
FT METAL 145 145 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY)
FT METAL 148 148 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY)
FT METAL 152 152 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY)
FT METAL 171 171 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY)
FT METAL 174 174 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY)
FT METAL 177 177 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY)
FT METAL 181 181 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY)
FT METAL 210 210 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY)
FT METAL 213 213 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY)
FT METAL 216 216 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY)
FT METAL 220 220 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY)
FT METAL 240 240 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY)
FT METAL 243 243 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY)
FT METAL 246 246 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY)
FT METAL 250 250 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY)
FT METAL 306 306 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY)
FT METAL 309 309 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY)
FT METAL 312 312 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY)
FT METAL 316 316 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY)
FT METAL 343 343 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY)
FT METAL 346 346 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY)
FT METAL 349 349 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY)
FT METAL 353 353 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY)
SQ SEQUENCE 394 AA: 43342 MW: 9C45BC4F CRC32:

```

```

Query Match 55.8% Score 63; DB 1; Length 394;
Best Local Similarity 53.8%; Pred. No. 1.88e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 83 CPICIKVDDP 95
Qy 3 CNIKVQVNDNF 15

```

```

RESULT 6 PRELIMINARY; PRT: 2606 AA.
ID Q21920 Q21920;
AC Q21920: Q21920;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE R11A8.7 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA BARDELL S.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALLDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RA COMINGS P.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: 270686; CA94615.1; -.
DR EMBL: 270310; CA94615.1; JOINED.
DR EMBL: 270310; CA94370.1; -.
DR EMBL: 270686; CA94370.1; JOINED.
DR HSSP: Q00421; LAMC.
DR PFM: PF00023; ank. 19.
DR PFM: PF00013; KH-domain 1.
SQ SEQUENCE 2606 AA: 285415 MW: FAD767A8 CRC32:

```

```

Query Match 55.8% Score 63; DB 5; Length 2606;
Best Local Similarity 53.8%; Pred. No. 1.88e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 965 CAISVRDMSNFP 977
Qy 3 CNIKVQVNDNF 15

```

```

RESULT 7 PRELIMINARY; PRT: 379 AA.
ID Q08356 Q08356;
AC Q08356: Q08356;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE DNAS PROTEIN.
OS Rhodospseudomonas sp.
OC Bacteria; Proteobacteria; alpha subdivision; Bradyrhizobium group;
OC Rhodospseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-NO. 7;
RX MEDLINE: 97214635.
RA MOMMA K., INUI M., YAMAGATA H., YUKAWA H.;
RT "Cloning of dnaK and dnaJ homologous genes from a purple non-sulfur
RT bacterium Rhodospseudomonas species."
RL Biochim. Biophys. Acta 150:235-239(1997).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAS, AND TO EUKARYOTIC DNAS-LIKE
CC PROTEINS.
DR EMBL: D78133; BAA19797.1; -.
DR HSSP: P08622; 1XBL.
DR PROSITE: PS00636; DNAS_1; 1.
DR PROSITE: PS00637; DNAS_CXXCXGXG; 1.
DR PFM: PF00226; Dnas; 1.
DR PFM: PF00684; Dnas_CXXCXGXG; 1.
DR PFM: PF01556; Dnas_C; 1.
DR PRINTS: PR00625; DNASPROTEIN.
KW Chapterone: DNA replication.
SQ SEQUENCE 379 AA: 40992 MW: 5D28C6EE CRC32:

```

```

Query Match 54.9% Score 62; DB 2; Length 379;
Best Local Similarity 38.5%; Pred. No. 3.06e-01;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

```

```

Db 44 QCEIKFEINEAY 56
Qy 2 ECNIKVDVNDNF 14

```

```

RESULT 8 PRELIMINARY; PRT: 501 AA.
ID Q02213 Q02213;
AC Q02213: Q02213;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PUTATIVE G PROTEIN-COUPLED RECEPTOR C02D4.2.
GN C02D4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

```

[1]
RN SEQUENCE FROM N.A.
RP BAYNES C.;
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PUTATIVE G-PROTEIN RECEPTOR.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC MOST SIMILAR TO INSECT OCTOPAMINE RECEPTORS.
DR EMBL: 281031; CAB02718.1; -.
DR WORMPEP: C02D4.2; CE07837.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR: 1.
DR PFM: PF00001; 7tm.1; 2.
DR PRITS: PR00237; GPCR_HODOPSIN.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 49
FT TRANSMEM 50 72
FT DOMAIN 73 82
FT TRANSMEM 83 104
FT TRANSMEM 120 136
FT TRANSMEM 137 157
FT TRANSMEM 158 177
FT TRANSMEM 178 200
FT TRANSMEM 201 224
FT TRANSMEM 225 246
FT TRANSMEM 247 373
FT TRANSMEM 374 395
FT TRANSMEM 396 426
FT TRANSMEM 427 447
FT TRANSMEM 448 487
FT TRANSMEM 493 496
FT CARBOHYD 15 15
FT DISULFID 134 212
SQ SEQUENCE 501 AA: 56404 MW: 18F1C57F CRC32;
BY SIMILARITY.
Query Match 54.9%; Score 62; DB 5; Length 501;
Best Local Similarity 54.5%; Pred. No. 3,06e-01;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
DB 336 CKVKVDVND 346
OY 3 CNIKVKVDVND 13
RESULT 9
ID 076322 PRELIMINARY: PRT: 466 AA.
AC 076322;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE SYNAPSIN S-SYN-SHORT (FRAGMENT).
OS Loligo pealeii (Longfin squid).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
CC Myposida; Loliginidae; Loligo.
RN [1]
RP SEQUENCE FROM N.A.
RA HILFICKER S., SCHWEITZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
RA AUGUSTINE G.J.;
RT "Two Sites of Action for Synapsin Domain E in Regulating
RT Neurotransmitter Release.";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055373; AAC24822.1; -.
DR HSSP: P17599; IAUx.
FT NON_TER 1
SQ SEQUENCE 466 AA: 52133 MW: C751B337 CRC32;
Query Match 53.1%; Score 60; DB 5; Length 466;
Best Local Similarity 35.7%; Pred. No. 7,98e-01;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
DB 287 CKCDIHYOKIGNNY 300
OY 1 CECNIKVKVDVNDNF 14

RESULT 10
ID 076323 PRELIMINARY: PRT: 503 AA.
AC 076323;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE SYNAPSIN S-SYN-LONG (FRAGMENT).
OS Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Teuthoidea;
CC Myposida; Loliginidae; Loligo.
RN [1]
RP SEQUENCE FROM N.A.
RA HILFICKER S., SCHWEITZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
RA AUGUSTINE G.J.;
RT "Two Sites of Action for Synapsin Domain E in Regulating
RT Neurotransmitter Release.";
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055374; AAC24823.1; -.
DR HSSP: P17599; IAUx.
FT NON_TER 1
SQ SEQUENCE 503 AA: 55738 MW: 693971D9 CRC32;
Query Match 53.1%; Score 60; DB 5; Length 503;
Best Local Similarity 35.7%; Pred. No. 7,98e-01;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
DB 287 CKCDIHYOKIGNNY 300
OY 1 CECNIKVKVDVNDNF 14
RESULT 11
ID 077704 PRELIMINARY: PRT: 814 AA.
AC 077704;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE DESMOCOLLIN TYPE 2 (FRAGMENT).
GN DSC2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC Eutheria; Carnivora; Flissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBERTS G.A., BURDETT I.D., PIDSLLEY S.C., KING I.A., MAGEE A.I.,
RA BUXTON R.S.;
RT "Antisense expression of a desmocollin gene in MDCK cells alters
RT desmosome plaque assembly but does not affect desmoglein expression.";
RL Eur. J. Cell Biol. 76:192-203(1998).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AJ002299; CAA05309.1; -.
DR HSSP: P09803; ISUH.
DR PROSITE: PS00232; CADHERIN: 3.
DR PFM: PF00028; cadherin: 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN 101 814
SQ SEQUENCE 814 AA: 91081 MW: BE0007A8 CRC32;
Query Match 53.1%; Score 60; DB 6; Length 814;
Best Local Similarity 61.5%; Pred. No. 7,98e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 306 CIINIDVNDLP 318
OY 3 CNIKVKVDVNDNF 15
RESULT 12
ID 09XU08 PRELIMINARY: PRT: 265 AA.
AC 09XU08;

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE T06G6.11 PROTEIN.
 GN T06G6.11.
 OS *Caenorhabditis elegans*.
 OC Euarvota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RX KERSHAM J.
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RX WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BOWFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DUREIN R., FAYELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIRI M., JOHNSTON L.,
 RA JONES M., KERSHAM J., KIRSTEN J., LAISTER N., LAUREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STRADEN J., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Nature 368:37-38(1994).
 SO EMBL; 281587; CAB04709.1; -.
 SO SEQUENCE 265 AA; 30553 MW; 8CC2E535 CRC32;

Query Match	52.2%	Score 59	DB 5	Length 265
Best Local Similarity	46.2%	Pred. NO.	1.26e+00	
Matches	6	Conservative	5	Mismatches 2
				Indels 0
				Gaps 0
Db	190 DCCWBMKDNPONF	202		
Oy	2 ECNIRKVDNDNF	14		

Query Match	52.28;	Score 59;	DB 14;	Length 659;
Best Local Similarity	53.38;	Pred. No. 1.28e+00;		
Matches 8;	Conservative	2;	Mismatches 5;	Indels 0;
				Caps 0;
Db 303	CENSIRKDIVGNCFP	317		
	: :			

QY 1 CECNIKVDVNDNFP 15

RESULT 14
ID Q9WT02 PRELIMINARY; PRT; 780 AA.
AC Q9WT02;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ORIGIN BINDING PROTEIN.
GN U73.
OS Human herpesvirus 6.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HST;
RA ISOGAWA Y., MUKAI T., NAKANO K., KAGAWA M., CHEN J., MORI Y.,
RA SUNAGAWA T., SASAHARA J., ZOU P., KOSUGE H., YAMANISHI K.;
RT "A comparison of the complete DNA sequences between human herpesvirus-
RT 6 variant A and B";
RL J. Virol. 0:0-0(1999).
DR EMBL, AB021506; BAA78294.1; -
SQ SEQUENCE 780 AA; 89540 MW; 5F3A240A CRC32;

Query Match 52.2%; Score 59; DB 14; Length 780;
Best Local Similarity 53.3%; Pred. No. 1.28e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 424 CENSIKVDVNGNFP 438
II::II::I::I
QY 1 CECNIKVDVNDNFP 15

RESULT 15
ID C001912 PRELIMINARY; PRT; 2163 AA.
AC C001912;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS.
GN R10F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditia; Rhabditidae; Rhabditidae; Pelodetidae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
RA JONES M., KERSHAM J., KIRSTEN J., LAISER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PEGGY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-IEG J., THOMAS K., VAUDIN M., VAUGHAN R., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA DU Z., GATTUNG S.;
RN Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RN Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
C -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL: AF003388; AAB54266.1; -
 DR HSSP: P15116; INCI.
 DR PROSITE: PS00232; CADHERIN; 8.
 DR PFAM: PF00001; 7tm_1; 1.
 DR PFAM: PF00028; cadherin; 15.
 DR PRINTS: PR00205; CADHERIN.
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 SQ SEQUENCE 2163 AA: 238609 MW: 6085652 CRC32;

Query Match 52.2%; Score 59; DB 5; Length 2163;
 Best Local Similarity 53.3%; Pred. No. 1.28e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1629 CKCHVYLDNDNLP 1643
 Oy 1 CECNKKVDNDNEP 15

Search completed: Sat May 13 07:21:16 2000
 Job time : 154 secs.

THIS PAGE BLANK (USPTO)

W08446 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch.p protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 07:29:32 2000; MasPar time 3.02 Seconds
Tabular output not generated. 117.801 Million cell updates/sec

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pap
Perfect Score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table: PAM 150
Gap 15

Searched: 188963 segs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:geneseqp

Statistics: Mean 19.232; Variance 58.683; scale 0.328

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	109	100.0	15	1	W08446 Self epitope of desmog	5.72e-05
2	109	100.0	614	1	Pemphigus vulgaris ant	5.72e-05
3	109	100.0	999	1	Human pemphigus vulgar	5.72e-05
4	56	51.4	480	1	Haem 84-2 portion of H	6.87e+01
5	56	51.4	516	1	Haem 84-1 portion of H	6.87e+01
6	55	50.5	431	1	Mouse CRF RBL receptor	8.71e+01
7	54	49.5	91	1	N-terminal region of b	1.10e+02
8	54	49.5	246	1	Protein able to bind t	1.10e+02
9	54	49.5	428	1	Chimeric KOMLV and Ty3	1.10e+02
10	54	49.5	429	1	Chimeric KOMLV and Ty3	1.10e+02
11	54	49.5	431	1	Rat CRF-beta receptor	1.10e+02
12	54	49.5	449	1	Chimeric KOMLV and Ty3	1.10e+02
13	54	49.5	1079	1	MuLV reverse transcript	1.10e+02
14	54	49.5	1196	1	Osteoinductive retrovi	1.10e+02
15	53	48.6	105	1	Variable light chain o	1.10e+02
16	53	48.6	105	1	Anti-human FasL antilo	1.10e+02
17	53	48.6	653	1	IDUA	1.40e+02
18	53	48.6	2115	1	Rubella virus RA27/3 N	1.40e+02
19	53	48.6	2205	1	Infectious rubella vir	1.40e+02
20	52	47.7	105	1	Antibody Lp1/2-6-3-VL	1.77e+02
21	52	47.7	105	1	Antibody Lp1/2-6-3-VL	1.77e+02
22	52	47.7	140	1	Human 5' EST secreted	1.77e+02
23	52	47.7	454	1	Kaposi's sarcoma assoc	1.77e+02

24	52	47.7	454	1	R97842	Kaposi's sarcoma assoc	1.77e+02
25	52	47.7	575	1	Y00157	Enterococcus faecalis	1.77e+02
26	52	47.7	601	1	Y00156	Enterococcus faecalis	1.77e+02
27	52	47.7	691	1	R04711	Sequence of guinea pig	1.77e+02
28	52	47.7	932	1	W97701	Staphylococcus aureus	1.77e+02
29	52	47.7	1844	1	W18302	Photobacterium luminesc	1.77e+02
30	52	47.7	1844	1	W18302	Photobacterium luminesc	1.77e+02
31	52	47.7	2504	1	W56557	Toxin TcdA, encoded by	1.77e+02
32	52	47.7	2504	1	W1871	Photobacterium luminesc	1.77e+02
33	51	46.8	39	1	R25101	bGRF proding analogue	2.23e+02
34	51	46.8	106	1	Y01034	Anti VLA-4 antibody HP	2.23e+02
35	51	46.8	106	1	W66742	Light chain variable r	2.23e+02
36	51	46.8	127	1	W06443	Humc3 VL region.	2.23e+02
37	51	46.8	128	1	W72435	Humanised VK sequence.	2.23e+02
38	51	46.8	128	1	W72433	Transplanted VK sequen	2.23e+02
39	51	46.8	128	1	R55211	Humanised HP1/2 light	2.23e+02
40	51	46.8	128	1	Y01036	Anti VLA-4 antibody SV	2.23e+02
41	51	46.8	128	1	R55209	Humanised HP1/2 kappa	2.23e+02
42	51	46.8	128	1	R59935	VK3 (DOWDY) VL.	2.23e+02
43	51	46.8	845	1	W56579	Fragment of toxin TcdA	2.23e+02
44	51	46.8	1261	1	W93601	Tomato Xa21 clone TRK2	2.23e+02
45	51	46.8	2516	1	W56572	Toxin TcdA, encoded by	2.23e+02

ALIGNMENTS

RESULT 1	W08446 standard; peptide: 15 AA.
ID	W08446; 18-FEB-1997 (first entry)
AC	Self epitope of desmoglein 3, implicated in autoimmune disease.
DE	Tolerisation: self-epitope: antigen: autoimmune disease.
KW	autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW	Pemphigus vulgaris; desmoglein; multiple sclerosis;
KW	herpes simplex virus; adenovirus; phosphonomonotase;
KW	human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW	influenza; haemagglutinin; reovirus; sigma protein.
OS	Homo sapiens.
ON	W09627387-A1.
RD	12-SEP-1996.
PR	07-MAR-1996; US03182.
PR	07-MAR-1995; US-400796.
PA	(HARD) HARVARD COLLEGE.
PI	Strominger JL, Wucherpfennig KW;
DR	WPT. 96-425218/42.
PT	Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT	antigens - useful in disease treatment, and method for
PT	identification of other self and non-self antigens implicated in
PT	auto-immune disease
PS	Claim 1: Page 41: 58pp: English.
CC	Pharmaceutical preparations for tolerisation to antigens comprise
CC	either an isolated human non-collagen or non-myosin basic protein
CC	(MPP) polypeptide which is capable of tolerising an individual to an
CC	autoantigen; or an isolated human pathogen polypeptide capable of
CC	tolerising an individual to that polypeptide. In both cases, the
CC	polypeptide (whether self or non-self) includes an amino acid
CC	sequence corresponding to a sequence motif for a MHC class II
CC	protein, such as HLA-DR, which is associated with a human autoimmune
CC	disease and which binds to the polypeptide to activate autoreactive
CC	T-cells in individuals with the autoimmune disease. This peptide is
CC	derived from the human desmoglein 3 protein (amino acids 512-526)
CC	and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC	derived from the human desmoglein protein are described in W08441-47.
S0	Sequence 15 AA:

Query Match 100.0%; Score 109; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 SARTLNRRYTGPTTF 15
Oy 1 SARTLNRRYTGPTTF 15

RESULT 2
ID W07908 standard; protein: 614 AA.
AC W07908:
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
dermatology.
OS Homo sapiens.
PN 308188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris autoantibody -
useful to treat and diagnose pemphigus vulgaris
PS Claim 1: Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 109; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 5,72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 511 SARTLNRRYTGPTF 525
QY 1 SARTLNRRYTGPTF 15

RESULT 3
ID R30742 standard; Protein: 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
N-PSDB: Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 109; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 5,72e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 512 SARTLNRRYTGPTF 526
QY 1 SARTLNRRYTGPTF 15

RESULT 4

ID R12099 standard; Protein: 480 AA.
AC R12099:
DT 22-JUL-1991 (first entry)
DE Haem 84-2 portion of H. contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN AU9062569-A.
PD 21-MAR-1991.
PF 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PR 01-MAR-1990; US-487181.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB: Q11799.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure; Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 480 AA;

Query Match 51.4%; Score 56; DB 1; Length 480;
Best Local Similarity 54.5%; Pred. No. 6,87e+01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 376 VNNRYGGRYK 386
QY 5 LNRRYTGPTF 15

RESULT 5
ID R12098 standard; Protein: 516 AA.
AC R12098:
DT 22-JUL-1991 (first entry)
DE Haem 84-1 portion of H. contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN AU9062569-A.
PD 21-MAR-1991.
PF 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PR 01-MAR-1990; US-487181.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB: Q11798.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure; Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 516 AA;

Query Match 51.4%; Score 56; DB 1; Length 516;
Best Local Similarity 54.5%; Pred. No. 6,87e+01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 484 VNNRYGGRYK 494
QY 5 LNRRYTGPTF 15

RESULT 6
ID R97293 standard; Protein: 431 AA.
AC R97293:
DT 21-AUG-1996 (first entry)
DE Mouse CRF RBL1 receptor.
KW Corticotropin releasing factor receptor; CRF-R; corticoliberin;

```

KW signal transduction.
OS Mus sp.
FH key Location/Qualifiers
FT modified_site 52
FT modified_site 61 /label= N-glycosylation_site
FT modified_site 61 /label= N-glycosylation_site
FT modified_site 94 /label= N-glycosylation_site
FT modified_site 105 /label= N-glycosylation_site
FT modified_site 113 /label= N-glycosylation_site
FT domain 139..159 /label= N-glycosylation_site
FT domain 169..188 /label= Transmembrane_domain-1
FT domain 196..229 /label= Transmembrane_domain-2
FT domain 245..265 /label= Transmembrane_domain-3
FT domain 285..307 /label= Transmembrane_domain-4
FT domain 331..351 /label= Transmembrane_domain-5
FT domain 365..385 /label= Transmembrane_domain-6
FT domain 365..385 /label= Transmembrane_domain-7
FT domain 365..385 /label= Transmembrane_domain-7
PN WO9617934-A2.
PD 13-JUN-1996.
PF 06-DEC-1995: U15909.
PR 09-DEC-1994: US-353537.
PR 17-JAN-1995: US-374009.
PA (SALK ) SALK INST BIOLOGICAL STUDIES.
PI Chen R, Donaldson CJ, Lewis KA, Perrin MR, Sawchenko P;
PI Vale MW; 96-287179/29.
DR N-PSDB: T28972.
PT Isolated corticotropin-releasing factor receptor (CRF-R) - used to
PT develop prods. for modulating signal transduction activity mediated
PT by CRF-R.
PS Claim 3; Page 83-85; 102pp; English.
CC Mouse corticotropin releasing factor receptor mCRF-RB1 was
CC identified as the product of a cDNA clone (T28972) isolated from a
CC mouse heart library. Recombinant mCRF-RB1 can be expressed in
CC host cells transformed by the cDNA clone. The receptor can be used
CC to identify agonists and antagonists that modulate the signal
CC transduction activity mediated by CRF receptors. It may be
CC administered therapeutically to reduce high ACTH levels caused by
CC excess CRF.
SQ Sequence 431 AA;

Query Match 50.5%; Score 55; DB 1; Length 431;
Best Local Similarity 46.2%; Pred. NO. 8.71e+01;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 47 RTTGNFSGPYTY 59
OY 3 RTLNRYTGPYTF 15

RESULT 7
ID P61040 standard; Protein: 91 AA.
AC P61040;
DE 03-JUL-1991 (first entry)
DE N-terminal region of bacterial expression vector.
KW Galactose: expression vector.
OS Synthetic.
PN J60248181-A.
PD 07-DEC-1985.
PF 23-MAY-1984: 102685.
PR 23-MAY-1984: JP-102685.
PA (SHTS ) SHITSIDO KK.
DR WPI: 86-025462/04.

```

```

DR N-PSDB: N60836.
PT Enzyme expression vector - giving transformant enzymes which can
PT be used as host enzymes in polypeptide biosynthesis.
PS Disclosure; Fig 4; 12pp; Japanese.
CC The vector may be used to express heterologous genes from a
CC transformed host when galactose is added as a carbon source.
SQ Sequence 91 AA;

Query Match 49.5%; Score 54; DB 1; Length 91;
Best Local Similarity 54.5%; Pred. NO. 1.10e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

DB 67 LNPRESYTYF 77
OY 5 LNNRYTGPYTF 15

RESULT 8
ID R40924 standard; Protein: 246 AA.
AC R40924;
DE 17-FEB-1994 (first entry)
DE Protein able to bind to HIV-1 tat protein.
KW Antibodies; antigen binding proteins; library; HIV;
KW Human Immunodeficiency Virus.
OS Synthetic.
PN EP-557897-A.
PD 01-SEP-1993.
PF 19-FEB-1993: 102609.
PR 28-FEB-1992: US-843125.
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
PI Dillon PJ, Rosen CA;
PI WPI: 93-274375/35.
DR N-PSDB: Q48605.
PT Antigen-binding proteins and corresp. synthetic genes - are
PT constructed synthetically and used to make antibody library, in
PT disease diagnosis, etc.
PS Claim 13; Page 22; 40pp; English.
CC The synthetic gene encodes a protein corresponding to an antibody
CC capable of binding to a specific antigen. In this case the HIV-1 tat
CC protein. Many synthetic genes are synthesised, each containing a
CC predetermined nucleotide region encoding the framework nucleotide
CC the heavy and light chains of antibody and underdetermined nucleotide
CC regions which are random sequences. The genes are then used in the
CC construction of vectors which are subsequently used to transform
CC microbes. The proteins thus produced are screened for binding
CC activity to the specific antigen.
SQ Sequence 246 AA;

Query Match 49.5%; Score 54; DB 1; Length 246;
Best Local Similarity 42.9%; Pred. NO. 1.10e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 57 ARYLDGSGNGRYTI 70
OY 2 ARTLNRYTGPYTF 15

RESULT 9
ID W08609 standard; Protein: 428 AA.
AC W08609;
DE 16-JUL-1997 (first entry)
DE Chimeric MoMuV and Ty3 integrase designated Atbmcm.
KW Moloney murine leukaemia virus; Saccharomyces cerevisiae;
KW retrotransposon; yeast Ty3; position specific integration;
KW inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KW Alzheimer's disease; rheumatoid arthritis; chimeric.
OS Chimeric - Moloney murine leukaemia virus.
OS Chimeric - Saccharomyces cerevisiae transposon Ty3.
FH key Location/Qualifiers
FT domain 1..61 /label= A
FT domain 62..283 /note= "From Ty3"

```

```

FT      /label-B
PI      /note-"From MOMLV"
FT      284. .428
FT      /label-C
FT      /note-"From MOMLV"
PN      WO637626-A1.
PD      28-NOV-1996.
PF      10-MAY-1996: U06727.
PR      22-MAY-1995: US-445466.
PA      (CHIR ) CHIRON VIAGENE INC.
PI      (REGC ) UNIV CALIFORNIA.
PI      Billachone VM, Dillidine SL, Jolly DJ, Respass JG;
PI      Sandmeyer SB;
PI      WPI: 97-021229/02.
PT      New chimeric integrase for targeted vector integration in
PT      eukaryotic genomes - useful for gene therapy, providing more
PT      consistent gene expression and lower rates of insertional
PT      mutagenesis
PS      Claim 7: Page -: 98pp: English.
CC      The present sequence is a specific example of a chimeric integrase
CC      made up of three domains from MOMLV and Ty3, with at least one domain
CC      derived from Ty3 integrase. This protein can direct integration of a
CC      vector construct into a defined region of a target eukaryotic genome.
CC      As part of gene delivery vehicles and transduction competent
CC      recombinant retroviral particles it can be useful in somatic
CC      and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC      genetic, infectious, degenerative, inflammatory, cardiovascular and
CC      autoimmune diseases or cancer. Typical examples of the many potential
CC      applications include treatment of HIV infection, haemophilia, Alzheimer's
CC      disease and rheumatoid arthritis. This protein rather than wild-type
CC      retroviral integrase reduces the rate of insertional mutagenesis and the
CC      degree of variation in gene expression, and particularly does not disrupt
CC      tRNA gene expression.
CC      N.B. The present sequence is not shown in the specification, but is
CC      made up from the two original sequences as specified, MOMLV and Ty3,
CC      which are shown.
SQ      Sequence 428 AA:

Query Match      49.5%: Score 54: DB 1: Length 428;
Best Local Similarity 46.2%: Pred. No. 1.10e+02;
Matches      6: Conservative      3: Mismatches      4: Indels      0: Gaps      0:

DB      360      TKNLEPRMKGPT 372
Oy      2      ARTLNRYTGPYT 14

RESULT 10
ID      WO8606 standard: protein: 429 AA.
AC      WO8606:
DT      16-JUL-1997 (first entry)
DE      Chimeric MOMLV and Ty3 integrase designated Ambtcm.
KW      Moloney murine leukaemia virus; Saccharomyces cerevisiae;
KW      retrotransposon; yeast Ty3; position specific integration;
KW      inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KW      Alzheimer's disease; rheumatoid arthritis; chimeric.
OS      Chimeric - Moloney murine leukaemia virus.
OS      Chimeric - Saccharomyces cerevisiae transposon Ty3.
FH      key
FH      location/Qualifiers
FT      1. .41
FT      /label-A
FT      /note-"From MOMLV"
FT      42. .284
FT      /label-B
FT      /note-"From Ty3"
FT      285. 429
FT      /label-C
FT      /note-"From MOMLV"
PN      WO637626-A1.
PD      28-NOV-1996.
PF      10-MAY-1996: U06727.
PR      22-MAY-1995: US-445466.
PA      (CHIR ) CHIRON VIAGENE INC.

```

```

PA      (REGC ) UNIV CALIFORNIA.
PI      Billachone VM, Dillidine SL, Jolly DJ, Respass JG;
PI      Sandmeyer SB;
PI      WPI: 97-021229/02.
PT      New chimeric integrase for targeted vector integration in
PT      eukaryotic genomes - useful for gene therapy, providing more
PT      consistent gene expression and lower rates of insertional
PT      mutagenesis
PS      Claim 7: Page -: 98pp: English.
CC      The present sequence is a specific example of a chimeric integrase
CC      made up of three domains from MOMLV and Ty3, with at least one domain
CC      derived from Ty3 integrase. This protein can direct integration of a
CC      vector construct into a defined region of a target eukaryotic genome.
CC      As part of gene delivery vehicles and transduction competent
CC      recombinant retroviral particles it can be useful in somatic
CC      and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC      genetic, infectious, degenerative, inflammatory, cardiovascular and
CC      autoimmune diseases or cancer. Typical examples of the many potential
CC      applications include treatment of HIV infection, haemophilia, Alzheimer's
CC      disease and rheumatoid arthritis. This protein rather than wild-type
CC      retroviral integrase reduces the rate of insertional mutagenesis and the
CC      degree of variation in gene expression, and particularly does not disrupt
CC      tRNA gene expression.
CC      N.B. The present sequence is not shown in the specification, but is
CC      made up from the two original sequences as specified, MOMLV and Ty3,
CC      which are shown.
SQ      Sequence 429 AA:

Query Match      49.5%: Score 54: DB 1: Length 429;
Best Local Similarity 46.2%: Pred. No. 1.10e+02;
Matches      6: Conservative      3: Mismatches      4: Indels      0: Gaps      0:

DB      361      TKNLEPRMKGPT 373
Oy      2      ARTLNRYTGPYT 14

RESULT 11
ID      R80575 standard: Protein: 431 AA.
AC      R80575;
DT      08-APR-1996 (first entry)
DE      Rat CRF2-beta receptor.
KW      CRF2-beta receptor; corticotropin-releasing factor-2 receptor;
KW      cerebrovascular disorder; memory disorder; Alzheimer disease.
OS      Rattus sp.
FH      key
FH      location/Qualifiers
FT      1. .117
FT      /label- Extracellular_N-terminal_domain
FT      118. .138
FT      /label- Transmembrane_domain
FT      139. .147
FT      /label- Intracellular_domain
FT      148. .167
FT      /label- Transmembrane_domain
FT      168. .184
FT      /label- Extracellular_domain
FT      185. .208
FT      /label- Transmembrane_domain
FT      209. .223
FT      /label- Intracellular_domain
FT      224. .244
FT      /label- Transmembrane_domain
FT      245. .261
FT      /label- Extracellular_domain
FT      262. .286
FT      /label- Transmembrane_domain
FT      287. .309
FT      /label- Intracellular_domain
FT      310. .329
FT      /label- Transmembrane_domain
FT      330. .342
FT      /label- Extracellular_domain
FT      343. .363

```

FT domain /label- Transmembrane domain
 FT 364. 431
 FT /label- C-terminal, intracellular domain
 PN W09534651-A2.
 PD 21-DEC-1995.
 PF 14-JUN-1995: U07757.
 PR 14-JUN-1995: US-259959.
 PR 31-JAN-1995: US-381433.
 PR 07-JUN-1995: US-485984.
 PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
 PI Chalmers D., De Souza EB., Grigoriadis DE., Liaw CW.
 PI Lovenberg TW., Oltersdorf T.
 DR MPI: 96-048680/05.
 DR N-ESDB: T12244.
 DR Corticotropin-releasing factor-2 receptor, and DNA encoding it -
 PT used to isolate CRF-2 receptor antagonists for the treatment of
 PT cerebrovascular disorders, memory disorders and Alzheimer's disease
 PS Disclosure: Page 63-66; 109pp: English.
 CC Rat corticotropin-releasing factor-2-beta (CRF2-beta) receptor
 CC (R90575) is a membrane-bound G-coupled protein receptor involved
 CC in signal transduction. It can be produced by expression of
 CC encoding cDNA (T12244) in prokaryotic or eucaryotic host cells.
 CC Recombinant CRF2 receptor is used to screen CRF2 receptor agonists
 CC and antagonists of therapeutic apptn., and to prepare antibodies
 CC which specifically bind to CRF2 receptors.
 SQ Sequence 431 AA:

Query Match 49.5%; Score 54; DB 1; Length 431;
 Best Local Similarity 38.5%; Pred. NO. 1.10e+02;
 Matches 5; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

Db 47 RTTRNFGSPYSY 59
 11 :|||:
 QY 3 RTLNRYTGPYT 15

RESULT 12
 ID W08608 standard; protein: 449 AA.
 AC W08608:
 DT 15-JUL-1997 (first entry)
 DE Chimeric MoMLV and Ty3 integrase designated AtbBcm.
 KM Moloney murine leukemia virus: Saccharomyces cerevisiae;
 KM retrotransposon; yeast Ty3; position specific integration;
 KM inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
 KM Alzheimer's disease; rheumatoid arthritis; chimeric.
 OS Chimeric - Moloney murine leukemia virus.
 OS Chimeric - Saccharomyces cerevisiae transposon Ty3.
 FH Key
 FH domain 1..61
 FT /label- A
 FT /note- "From Ty3"
 FT domain 62..304
 FT /label- B
 FT /note- "From Ty3"
 FT domain 305..449
 FT /label- C
 FT /note- "From MoMLV"
 PN W09637626-A1.
 PD 28-NOV-1996.
 PF 10-MAY-1996: U06727.
 PR 22-MAY-1995: US-445466.
 PA (CHIR) CHIRON VIAGENE INC.
 PA (REGC) UNIV CALIFORNIA.
 PI Blachone VM., Dildine SL., Jolly DJ., Respass JG.
 PI Sandmeyer SB.
 DR MPI: 97-021229/02.
 DR New chimeric integrase for targeted vector integration in
 PT eukaryotic genomes - useful for gene therapy, providing more
 PT consistent gene expression and lower rates of insertional
 PT mutagenesis
 PS Claim 7: Page -: 98pp: English.
 CC The present sequence is a specific example of a chimeric integrase
 CC made up of three domains from MoMLV and Ty3, with at least one domain

CC derived from Ty3 integrase. This protein can direct integration of a
 CC vector construct into a defined region of a target eukaryotic genome.
 CC As part of gene delivery vehicles and transduction competent
 CC recombinant retroviral particles it can be useful in somatic
 CC and germ cell gene therapy (in vivo or ex vivo) of a wide range of
 CC genetic, infectious, degenerative, inflammatory, cardiovascular and
 CC autoimmune diseases or cancer. Typical examples of the many potential
 CC applications include treatment of HIV infection, hemophilia, Alzheimer's
 CC disease and rheumatoid arthritis. This protein rather than wild-type
 CC retroviral integrase reduces the rate of insertional mutagenesis and the
 CC degree of variation in gene expression, and particularly does not disrupt
 CC tRNA gene expression.
 CC N.B. The present sequence is not shown in the specification, but is
 CC made up from the two original sequences as specified, MoMLV and Ty3,
 CC which are shown.
 SQ Sequence 449 AA:

Query Match 49.5%; Score 54; DB 1; Length 449;
 Best Local Similarity 46.2%; Pred. NO. 1.10e+02;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 381 TKNLPRMKGPYT 393
 2 ARTLNRYTGPYT 14

RESULT 13
 ID Y03163 standard; protein: 1079 AA.
 AC Y03163:
 DT 10-JUN-1999 (first entry)
 DE MuLV reverse transcriptase.
 KM MuLV reverse transcriptase: dsCDNA production: cDNA library production.
 OS Murine leukemia virus.
 OS US5891637-A.
 PN 06-APR-1999.
 PD 15-SEP-1997: 929967.
 PF 15-OCT-1996: US-732861.
 PR 15-SEP-1997: US-929967.
 PA (GETH) GENENTECH INC.
 PI Ruppert SJ.
 DR MPI: 99-253852/21.
 DT Producing double stranded cDNA molecules from mRNA transcripts and a
 PT novel xenotropic murine leukemia virus reverse transcriptase -
 PT useful for constructing cDNA libraries
 PS Claim 32: Column 45-52; 70pp: English.
 CC This sequence represents a murine leukemia virus (MuLV) reverse
 CC transcriptase. The invention relates to a method of producing double
 CC stranded cDNA (dsCDNA) molecules from mRNA transcripts and a xenotropic
 CC murine leukemia virus reverse transcriptase (XN-MuLV). The method, the
 CC host cells, reverse transcriptases and nucleic acids disclosed may be
 CC used to produce high quality, normalized, full-length, directionally
 CC cloned, cDNA libraries. Use of the method to construct cDNA libraries
 CC ensures that the reverse transcriptase is able extend the first strand of
 CC the cDNA up to the ultimate 5' end of the mRNA, therefore ensuring that
 CC all sequences are represented in the library, whereas previously, some
 CC sequences would be lost due to incomplete transcription (e.g. as a result
 CC of hairpin loop formation) or mRNA degradation (e.g. as a result of
 CC RNase H activity). Also, carrying the process out in cells means that enzymes
 CC and proteins present in the cell will repair any mistakes or nicks in the
 CC product DNA. Modifying the mRNA for insertion into a vector by adding the
 CC 5' cap provides a convenient way of priming it for cDNA synthesis. Also,
 CC the method provides a fast and clean way of synthesizing cDNA
 CC clones.
 SQ Sequence 1079 AA:

Query Match 49.5%; Score 54; DB 1; Length 1079;
 Best Local Similarity 46.2%; Pred. NO. 1.10e+02;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1011 TKNLPRMKGPYT 1023
 2 ARTLNRYTGPYT 14

RESULT 14
 ID R75189 standard; Protein; 1196 AA.
 AC R75189;
 DT 30-MAY-1996 (first entry)
 DE Osteoinductive retrovirus RFB-14 pol gene product.
 KW RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;
 KW bone development; osteoporosis; gene therapy; polymerase;
 KW reverse transcriptase.
 OS Retrovirus RFB-14.
 PN DE4411718-A1.
 PD 12-OCT-1995.
 PF 05-APR-1994: 411718.
 PR 05-APR-1994: DE-411718
 PA (GSFU-) GSF FORSCHUNGSZENTRUM UNIKELT & GESUNDHEIT.
 PI Erfle V, Gimbel W, Oestergaard M, Pedersen FS, Pedersen L;
 PI Schmidt J, Straus P;
 DR WPI: 95-352078/46.
 DR N-PSDB: 094266.
 PT RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins
 PS Clalm 14; Fig 1; 46pp; German.
 CC The full-length proviral genomic sequence of retrovirus RFB-14 has
 CC been determined. The virus codes for an osteoinductive protein,
 CC although the precise location of the coding region has not yet been
 CC identified. The virus may be useful in gene therapy of bone growth
 CC disorders such as osteoporosis. The present sequence is that of the
 CC viral pol gene product.
 SQ Sequence 1196 AA;

Query Match 49.5%; Score 54; DB 1; Length 1196;
 Best Local Similarity 46.2%; Pred. No. 1.10e+02;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1131 TKNEPRMKPYT 1143
 : : : : :
 OY 2 ARTLNRYTGYPT 14

RESULT 15
 ID W00832 standard; Protein; 105 AA.
 DT 29-MAY-1997 (first entry)
 DE Variable light chain of anti-human Fas ligand antibody NOK-5.
 KW Variable region; light chain; human; Fas ligand; monoclonal;
 KW antibody; NOK-5; hybridoma; inhibition; apoptosis; assay;
 KW diagnosis; disease; hepatitis; infectious mononucleosis;
 KW systemic lupus erythematosus.
 OS Mus musculus.
 PN W09629350-A1.
 PD 26-SEP-1996.
 PF 21-MAR-1996: J00734.
 PR 20-MAR-1995: JP-087420.
 PR 27-OCT-1995: JP-303492.
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 PI Kayagaki N, Nakata M, Okumura K, Yagita H;
 PI WPI: 96-443140/44.
 DR N-PSDB: T39558.
 PT Monoclonal antibody specifically recognising the Fas ligand - useful
 PT for the detection of Fas ligands either on cell surface or in
 PT solution
 PS Clalm 28; Page 91; 133pp; Japanese.
 CC The present sequence is the light chain variable region of the
 CC anti-human Fas ligand monoclonal antibody (Mab) NOK-5. NOK-5 is
 CC produced by the hybridoma NOK-5 (FERM BP-5048), which was prepared
 CC by immunising mice with transformed human Fas ligand expressing COS
 CC cells, and fusing spleen cells isolated from the mice with myeloma
 CC P3x63Ag8.653 (ATCC CRL-1580) cells. The Mab recognises the human
 CC Fas ligand on the cell surface or in solution, and can be used to
 CC inhibit the apoptosis inducing cell surface Fas ligand/Fas
 CC reaction. The Mab can also be used for a Fas ligand assay in
 CC biological samples (e.g. human blood), especially for disease
 CC diagnosis, e.g. hepatitis, infectious mononucleosis and systemic
 CC lupus erythematosus.

SQ Sequence 105 AA;

Query Match 48.6%; Score 53; DB 1; Length 105;
 Best Local Similarity 62.5%; Pred. No. 1.40e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 91 HYSSPYTF 98
 : : : : :
 OY 8 RYTGPTTF 15

Search completed: Sat May 13 07:29:39 2000
 Job time : 7 secs.

 WIDEORH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:22:27 2000: MasPar time 96.51 Seconds
 4.733 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pap
 Perfect Score: 109
 Sequence: 1 SARTLNRRYGPYTF 15

Scoring table:
 Gap 15
 PAM 150

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 27.244: Variance 33.691: scale 0.809

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	109	100.0	999	1	DSG3_HUMAN	8.31e-13
2	65	59.6	431	1	IMA4_YEAST	5.59e-02
3	62	56.9	391	1	GABG_XENLA	2.41e-01
4	60	55.0	273	1	HMO1_CAEEL	6.23e-01
5	60	55.0	334	1	VG08_BPT4	6.23e-01
6	59	54.1	304	1	CBP2_SIMVI	9.94e-01
7	58	53.2	444	1	NIFN_ANASP	1.58e+00
8	58	53.2	469	1	NIFD_METTH	1.58e+00
9	58	53.2	1237	1	POLA_DROME	1.58e+00
10	57	52.3	387	1	GAT6_CHICK	2.49e+00
11	57	52.3	1013	1	CHSA_EMENT	2.49e+00
12	56	51.4	360	1	IDH1_YEAST	3.90e+00
13	56	51.4	1317	1	NUCLOPORIN NP145 (NU	3.90e+00
14	55	50.5	245	1	GVPL_ANAFI	6.08e+00
15	55	50.5	261	1	Y4ZC_RHISN	6.08e+00
16	55	50.5	391	1	GABG_XENLA	6.08e+00
17	55	50.5	431	1	CRF2_MOUSE	6.08e+00
18	55	50.5	433	1	CBPA_ANOGA	6.08e+00
19	55	50.5	455	1	KR2_EBV	6.08e+00
20	54	49.5	120	1	POL1_MOUSE	9.41e+00
21	54	49.5	130	1	V29K_TRVTC	9.41e+00
22	54	49.5	194	1	VNA3_VACCV	9.41e+00
23	54	49.5	231	1	RECO_COXBU	9.41e+00

24	54	49.5	252	1	V29K_TRVSY	29 KD PROTEIN.	9.41e+00
25	54	49.5	282	1	POL_MLYCB	POL POLYPROTEIN [CONTA	9.41e+00
26	54	49.5	365	1	GAL7_YEAST	GALACTOSE-1-PHOSPHATE	9.41e+00
27	54	49.5	367	1	V669_METJA	PUTATIVE ATP-DEPENDENT	9.41e+00
28	54	49.5	370	1	GPRA_RAT	PROBABLE G PROTEIN-COU	9.41e+00
29	54	49.5	390	1	POL3_MOUSE	RETROVIRUS-RELATED POL	9.41e+00
30	54	49.5	428	1	YUPL_YEAST	PROBABLE MANNOSYLTRANS	9.41e+00
31	54	49.5	581	1	POL_MLYRK	POL POLYPROTEIN [CONTA	9.41e+00
32	54	49.5	746	1	YLS8_CAEEL	HYPOTHETICAL 83.6 KD P	9.41e+00
33	54	49.5	843	1	POL_MLYAK	POL POLYPROTEIN [CONTA	9.41e+00
34	54	49.5	1196	1	POL_MLYVD	POL POLYPROTEIN [CONTA	9.41e+00
35	54	49.5	1196	1	POL_MLYAV	POL POLYPROTEIN [CONTA	9.41e+00
36	54	49.5	1199	1	POL_MLYVO	POL POLYPROTEIN [CONTA	9.41e+00
37	54	49.5	1204	1	POL_MLYVF	POL POLYPROTEIN [CONTA	9.41e+00
38	54	49.5	1204	1	POL_MLYV5	POL POLYPROTEIN [CONTA	9.41e+00
39	54	49.5	1204	1	POL_MLYVF	POL POLYPROTEIN [CONTA	9.41e+00
40	53	48.6	109	1	Y85A_METJA	HYPOTHETICAL PROTEIN M	1.45e+01
41	53	48.6	228	1	Y85C_EMENT	HYPOTHETICAL 25.4 KD P	1.45e+01
42	53	48.6	245	1	YFIO_ECOLI	HYPOTHETICAL 27.8 KD L	1.45e+01
43	53	48.6	2205	1	POLN_RUBVT	NONSTRUCTURAL POLYPROT	1.45e+01
44	53	48.6	2261	1	RRPL_MUMPM	RNA POLYMERASE BETA SU	1.45e+01
45	53	48.6	2769	1	THYG_BOVIN	THYOGLOBULIN PRECURSO	1.45e+01

ALIGNMENTS

RESULT 1
 ID DSG3_HUMAN STANDARD: PRT: 999 AA.

AC P32926;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).
 GN DSG3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP MEDLINE: 92069753.
 RA Amagai M., Klaus-Kovtun V., Stanley J.R.;
 RT "Autoantibodies against a novel epithelial cadherin in pemphigus
 vulgaris, a disease of cell adhesion.";
 RL Cell 67:869-877(1991).
 CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND
 CC CARCINOMAS.
 CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -!- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN
 CC DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE
 CC LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES
 CC AGAINST DSG3.
 CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M76482; AAA60230.1; -;
 CC PIR: A41088; IJHUG3.
 CC HSSP: P09803; IEDH.
 CC MIM: 169655; -;
 CC PROSITE: PS00232; CADHERIN. 3.
 CC PFMW: PF00028; cadherin. 4.
 CC Cell adhesion: Signal; Transmembrane; Cytoskeleton; Glycoprotein;

KW Calcium-Binding: Repeat.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 30 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA: 107503 MW: 60479DD46AC219A1 CRC64;
 Query Match 100.0% Score 109; DB 1; Length 999;
 Best Local Similarity 100.0% Pred. No. 8,31e-13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 SARTLNRYTGPYTF 526
 1 SARTLNRYTGPYTF 15

RESULT 2
 ID IM44_YEAST STANDARD: PRT: 431 AA.
 AC 001852;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
 DE PRECURSOR (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 1) (INNER MEMBRANE
 DE IMPORT SITE PROTEIN 45) (ISPA5) (MEMBRANE IMPORT MACHINERY PROTEIN
 DE IM44).
 GN TIM44 OR MPI1 OR MIM44 OR ISPA5 OR YIL022M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB320;
 RX MEDLINE: 93010953.
 RA Mease A.C., Blom J., Griwell L.A., Melzer M.;
 RT "MPI1, an essential gene encoding a mitochondrial membrane protein,
 RT is possibly involved in protein import into yeast mitochondria.";
 RL EMO J. 11:3619-3628(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Guelles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Welsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE: 93345448.
 RA Horst M., Jenoe P., Kronidou N.G., Bolliger L., Oppliger W.,
 RA Scherer P., Manning-Krieg U., Jascut T., Schatz G.;
 RT "Protein import into yeast mitochondria: the inner membrane import
 RT site protein ISPA5 is the MPI1 gene product.";
 RL EMO J. 12:3035-3041(1993).
 CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
 CC -1- PROBABLY INVOLVED IN TRANSLLOCATION ACROSS THE INNER MEMBRANE.
 CC AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
 CC PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE
 CC (MGE1) TO DRIVE PROTEIN TRANSLLOCATION INTO THE MATRIX USING ATP

CC AS AN ENERGY SOURCE.
 CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
 CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 246881, CA86970.1; -;
 CC EMBL: X67276; CA47693.1; -;
 CC PIR: S25196; S25196.
 CC SCD: L0001138; MPI1.
 DR Mitochondrion; Inner membrane; Transport; Protein transport;
 KW Translocation; Transit peptide; ATP-binding.
 FT TRANSIT 1 431 MITOCHONDRION;
 FT CHAIN ? 431 MITOCHONDRIAL IMPORT INNER MEMBRANE
 FT NP_BIND ? 108 TRANSLOCASE SUBUNIT TIM44.
 FT SEQUENCE 431 AA: 48854 MW: 8698C738178B1E44 CRC64;
 Query Match 59.6% Score 65; DB 1; Length 431;
 Best Local Similarity 57.1% Pred. No. 5.39e-02;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 13 SSRTIARYRSQYT 26
 1 SARTLNRYTGPYTF 14

RESULT 3
 ID GAB6_XENLA STANDARD: PRT: 391 AA.
 AC P70005;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TRANSCRIPTION FACTOR GATA-6B (GATA BINDING FACTOR-6B).
 DE GATA-6A OR GATA-6.
 GN Xenopus laevis (African clawed frog).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 97180937.
 RA Gove C.D., Wamsley M., Nijjar S., Bertwistle D., Gullie M.,
 RA Parlington G., Bomford A., Patient R.;
 RT "Over-expression of GATA-6 in Xenopus embryos blocks differentiation
 RT of heart precursors.";
 RL EMO J. 16:355-368(1997).
 CC -1- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE
 CC CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES
 CC THE EXPRESSION OF CARDIAC MYC-ALPHA IN VIVO.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y08865; CA470088.1; -;
 CC HSP: P1678; 1GAU
 CC PRINTS: PR00619; GATA2NFINGER.

DR PROSITE: PS00344: GATA_ZN_FINGER. 2.
DR PFAM: PF00320: GATA. 2.
KM Transcription regulation: Activator: DNA-binding; Zinc-finger:
Nuclear protein.
FT ZN_FING 182 206 GATA-TYPE.
FT ZN_FING 236 260 GATA-TYPE.
FT DOMAIN 70 79 POLY-SER.
FT DOMAIN 241 245 POLY-THR.
FT DOMAIN 293 298 POLY-SER.
SQ SEQUENCE 391 AA: 41503 MW: C93C0CDE246204D3 CRC64:
Query Match 56.9%: Score 62; DB 1; Length 391;
Best Local Similarity 61.5%: Pred. No. 2,41e-01;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 120 ARPLNGSYGSPYT 132
QY 2 ARTLNRRYTGPT 14
RESULT 4
ID HMD1_CABEL STANDARD: PRT: 273 AA.
AC Q18273:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE HOMEBOX PROTEIN C28A5.4.
GN C28A5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemata; Secernentea; Rhabdilia; Rhabdilitida;
OC Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
RN [1]
RP STRAIN-BRISTOL N2;
RA Palmer S.;
RL Submitted (APR-1994) to the EMBL/Genbank/DBJ databases.
CC -1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1 SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z32680: CAA83601.1; -;
DR HSSP: P02836: 1HD.
DR WORMPEP: C28A5.4; CE05325.
DR PRINTS: PR00024: HOMEBOX.
DR PRINTS: PR00031: HTHREPRESSR.
DR PROSITE: PS00027: HOMEBOX_1; 1.
DR PROSITE: PS00071: HOMEBOX_2; 1.
DR PFAM: PF00046: homeobox; 1.
KW Hypothetical protein: Homeobox: DNA-binding; Nuclear protein.
FT DNA_BIND 102 161 HOMEBOX.
SQ SEQUENCE 273 AA: 30163 MW: 1B933F08C2576EB CRC64:
Query Match 55.0%: Score 60; DB 1; Length 273;
Best Local Similarity 40.0%: Pred. No. 6,23e-01;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Db 215 PPMTEQYASPYLY 229
QY 1 SARTLNRRYTGPT 15
RESULT 5
ID VGO8_BPT4 STANDARD: PRT: 334 AA.
AC P19062:
DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE BASEPLATE STRUCTURAL PROTEIN GP8.
GN 8.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D;
RX MEDLINE: 90384864.
RA Efimov V.P., Prilipov A.G., Mesyanzhinov V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 6, 7 and 8."
RL Nucleic Acids Res. 18:5313-5315(1990).
CC -1 FUNCTION: STRUCTURAL COMPONENT OF THE BASEPLATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X15907: CAA34023.1; -;
DR PIR: J00658: G8BPT4.
DR PIR: JF0058: JF0058.
KW Structural protein.
SQ SEQUENCE 334 AA: 38008 MW: 4997860773E14899 CRC64:
Query Match 55.0%: Score 60; DB 1; Length 334;
Best Local Similarity 60.0%: Pred. No. 6,23e-01;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 97 DTRYPDPYTF 106
QY 6 NNRYTGPT 15
RESULT 6
ID CBP2_SIMV1 STANDARD: PRT: 304 AA.
AC P42788:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).
OS Simulium vittatum (Black fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Simuliidae; Simulium.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-GUT;
RX MEDLINE: 94093864.
RA Ramos A., Mahowald A., Jacobs-Lorena M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins".
RL Insect Mol. Biol. 1:149-163(1993).
CC -1 FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1 TISSUE SPECIFICITY: GUT-SPECIFIC.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L08481: AAA18531.1; -;
DR HSSP: P00730: 4CPA.

DR PROSITE: PS00132: CARBOXYPEPT_2N.1; 1.
 DR PROSITE: PS00133: CARBOXYPEPT_2N.2; 1.
 DR PFAM: PF00246: Zn_carboxypept. 1.
 KW Hydrolase: Carboxypeptidase; Zinc.
 FT METAL 58 58 ZINC (BY SIMILARITY).
 FT METAL 61 61 ZINC (BY SIMILARITY).
 FT METAL 184 184 ZINC (BY SIMILARITY).
 FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 125 148 BY SIMILARITY.
 SO SEQUENCE 304 AA: 34849 MW: 26633FBA6A9144 CRC64;
 Query Match 54.1%; Score 59; DB 1; Length 304;
 Best Local Similarity 50.0%; Pred. No. 9.94e-01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 214 AARSLAORVGTDT 227
 :||:| || ||
 Oy 1 SARTLNRRYGTPT 14

RESULT 7
 ID NIFN_AMASP STANDARD; PRT; 444 AA.
 AC 044145;
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFN.
 GN NIFN.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria: Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 RN (1)
 RP Sequence from N.A.
 RA Birkema W.J., Scarpino L.A., Haselkorn R.
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
 CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
 CC -!- PATHWAY: FEMO COFACTOR BIOSYNTHESIS.
 CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL: U47055; AAA87948.1; -;
 DR PROSITE: PS00699; NITROGENASE_1_1; 1.
 DR PROSITE: PS00090; NITROGENASE_1_2; FALSE_NEG.
 DR PFAM: PF00148; oxidored_nitro. 1.
 KW Nitrogen fixation.
 SO SEQUENCE 444 AA: 48332 MW: 7668486165959037 CRC64;
 Query Match 53.2%; Score 58; DB 1; Length 444;
 Best Local Similarity 46.2%; Pred. No. 1.58e+00;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 245 AARLEDRFNIPY 257
 :||:| || ||
 Oy 1 SARTLNRRYGTPT 13

RESULT 8
 ID NIFD_METH STANDARD; PRT; 469 AA.
 AC 027605;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NITROGENASE MOLYBDENUM-IRON PROTEIN ALPHA CHAIN (EC 1.18.6.1)
 DE (NITROGENASE COMPONENT 1) (DINITROGENASE).

GN NIFD OR MTH1563.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea: Euryarchaeota: Methanobacteriales: Methanobacteriaceae;
 OC Methanobacterium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H.
 RX MEDLINE: 98037514.
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-W., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patel D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-T., Rice P., Nollig J., Reeve J.N.:
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT delta: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE
 CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
 CC -!- CATALYTIC ACTIVITY: 8 REDUCED FERREDOXIN + 8 H⁺ + N₂ + 16 ATP
 CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS
 CC 30-32 FE, 2 MO. AND INORGANIC SULFUR.
 CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL: AE000916; AAB86037.1; -;
 DR HSSP: P00467; IMIO.
 DR PROSITE: PS00699; NITROGENASE_1_1; FALSE_NEG.
 DR PROSITE: PS00090; NITROGENASE_1_2; 1.
 DR PFAM: PF00148; oxidored_nitro. 1.
 KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur.
 SO SEQUENCE 469 AA: 52866 MW: 7D2CD5FEDE9EC7B8 CRC64;
 Query Match 53.2%; Score 58; DB 1; Length 469;
 Best Local Similarity 58.3%; Pred. No. 1.58e+00;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 273 ARLLEERYNVY 284
 :||:| || ||
 Oy 2 ARLTNRRYGTPT 13

RESULT 9
 ID POLA_DROME STANDARD; PRT; 1237 AA.
 AC P10394;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RETROVIRUS-RELATED POLYPROTEIN FROM TRANSPOSON 412 [CONTAINS:
 DE PROTEASE (EC 3.4.23.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
 DE ENDONUCLEASE].
 GN POL.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86274717.
 RA Yuki S., Inouye S., Ishimaru S., Saigo K.:
 RT "Nucleotide sequence characterization of a Drosophila
 RT retrolansposon, 412.";
 RL Eur. J. Biochem. 158:403-410(1986).


```

Db      16  ARTASHPERYSY 29
      111 ::::1 :::
Oy      2  ARTLNRRITGPTTF 15

RESULT 12
ID      IDH1 YEAST          STANDARD:          PRT:          360 AA.
AC      P28834.
DT      01-DEC-1992 (Rel. 24, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      ISOCITRATE DEHYDROGENASE [NAD] SUBUNIT 1, MITOCHONDRIAL PRECURSOR
DE      (EC 1.1.1.41) (ISOCITRIC DEHYDROGENASE) (NAD+-SPECIFIC ICSDH).
GN      IDH1 OR YNL037C OR N2690.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC      Saccharomycetaceae; Saccharomyces.
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 49-61; 72-83; 325-333 AND 359-356.
RA      MEDLINE: 92355609.
RX      CUPP J.R., McAllister-Henn L.;
RT      *Cloning and characterization of the gene encoding the IDH1 subunit
RT      of NAD(+) dependent isocitrate dehydrogenase from saccharomyces
RT      cerevisiae.*
RL      J. Biol. Chem. 267:16417-16423(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Duesthoeft A., Floeth M., Filtz C., Heuss-Netzel D.,
RA      Hilbert H., Moestl D.;
RA      Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 12-27.
RC      STRAIN-SC7.
RX      MEDLINE: 90330530.
RA      KEYS D.A., McAllister-Henn L.;
RT      *Subunit structure, expression, and function of NAD(H)-specific
RT      isocitrate dehydrogenase in Saccharomyces cerevisiae.*;
RL      J. Bacteriol. 172:4280-4287(1990).
RN      [4]
RP      RNA-BINDING.
RA      MEDLINE: 940899379.
RA      Elzinga S.D.J., Bednarz A.L., van Oosterum K., Dekker P.J.T.,
RA      Grivell L.A.;
RT      *Yeast mitochondrial NAD(+) dependent isocitrate dehydrogenase is an
RT      RNA-binding protein.*;
RL      Nucleic Acids Res. 21:5328-5331(1993).
CC      -1- FUNCTION: PERFORMS AN ESSENTIAL ROLE IN THE OXIDATIVE FUNCTION OF
CC      THE CITRIC ACID CYCLE. ALSO BINDS RNA; SPECIFICALLY TO THE
CC      5'-UNTRANSLATED LEADERS OF MITOCHONDRIAL MRNAs.
CC      -1- CATALYTIC ACTIVITY: ISOCITRATE + NAD(+) = 2-OXOGLOUTARATE +
CC      CO(2) + NADH.
CC      -1- ENZYME REGULATION: ALLOSTERICALLY REGULATED BY SEVERAL COMPOUNDS
CC      INCLUDING AMP, NAD+, AND CITRATE.
CC      -1- SUBUNIT: OCTAMER OF TWO NONIDENTICAL SUBUNITS IDH1 AND IDH2.
CC      -1- SUBCELLULAR LOCATION: MITOCHONDRIAL
CC      -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLALATE
CC      DEHYDROGENASES FAMILY.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: M95203; AAA34711.1; -
DR      EMBL: J71313; CA95904.1; -
DR      PIR: S31264; S31264.
DR      HSSP: P00351; 10SI.
DR      SCD: L0000849; IDH1.
DR      PROSITE: PS00470; IDH_1MDH; 1.

```

```

DR PFAM: PF00180; Isohdn. 1.
KW Oxidoreductase; NAD; Tricarboxylic acid cycle; Transl. peptide;
KW Mitochondrion; Allosteric enzyme; RNA-binding.
FT TRANSIT 1 11 MITOCHONDRON
FT CHAIN 12 360 ISOCITRATE DEHYDROGENASE (NAD) SUBUNIT 1.
FT ACT_SITE 103 103 BINDING TO ISOCITRATE (BY SIMILARITY).
SQ SEQUENCE 360 AA; 39324 MM; 0932E7B3CD685240 CRC64;

Query Match 51.4%; Score 56; DB 1; Length 360;
Best Local Similarity 42.9%; Pred. NO. 3,90E+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 17 AERTLPKRYGGRFT 30
: |||::|:|
Oy 1 SARTLNNRRTGPGYT 14

RESULT 13
ID N145_YEAST STANDARD: PRT: 1317 AA.
AC P49687;
DT 01-FEB-1996 (Rel. 33, Created)
DI 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 33, Last annotation update)
DE NUCLEOPORIN NUP145 (NUCLEAR PORE PROTEIN NUP145).
GN NUP145 OR YGLO92W.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94320139.
RA Fader E., Boelems W.C., Wimmer C., Mattei I.W., Hurt E.C.:
RT "Nup145p is required for nuclear export of mRNA and binds
RT homopolymeric RNA in vitro via a novel conserved motif."
RL Cell 78:275-289(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 26109 / X2180;
RX MEDLINE: 94253246.
RA Wentle S.R., Blobel G.:
RT "NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine
RT (GLFEG) nucleoporin required for nuclear envelope structure."
RL J. Cell Biol. 125:955-969(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE: 97435481.
RA Rieger M., Bruechner M., Schaefer M., Mueller-Auer S.:
RT "Sequence analysis of 203 klobases from Saccharomycetes cerevisiae
RT chromosome VII".
RL Yeast 13:1077-1090(1997).
CC -1- FUNCTION: MAY PLAY A DIRECT ROLE IN NUCLEOCYTOPLASMIC EXPORT OF
CC RNA AND ALSO IN PROTEIN IMPORT. BINDS HOMOPOLYMERIC RNA.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X76557; CAAS4057.1; -
DR EMBL: Z32672; CAAB3584.1; -
DR EMBL: Z72614; CAAG6798.1; -
DR SGD: L0001294; NUP145.
KW Nuclear protein; Transport; Repeat; RNA-binding.
CC CONFLICT 281 282 NA -> OR (IN REF. 2).
FT CONFLICT 1142 1142 L -> S (IN REF. 2).
FT CONFLICT 1310 1317 LMCTYRI -> FEVYI (IN REF. 2).

```

SO SEQUENCE 1317 AA: 145660 MW: 59399086B553030 CRC64:

Query Match 51.4% Score 56; DB 1; Length 1317;
Best Local Similarity 46.2%; Pred. No. 3.90e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1145 NAOALMDRYGNY 1157
1 SARTLNRYTGPY 13

RESULT 14
ID GVPL ANAFL STANDARD; PRT; 245 AA.
AC P55149;

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 35, Last annotation update)
DE GVPE/L PROTEIN.
GN GVPEL.

OS Anabaena flos-aquae.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.

RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN-CCAP 1403/13F;
RX MEDLINE: 97217685.

RA Kinsman R., Hayes P.K.;
RT Genes encoding proteins homologous to halobacterial Gyps N, J, K, F
RT flos-aquae.
RT DNA Seq. 7:97-106(1997).
CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS
CC VESICLE SYNTHESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U17109; AAA58714.1; -.
KW Gas vesicle.

SO SEQUENCE 245 AA: 28676 MW: 5CF0BFC999E11BC0 CRC64;

Query Match 50.5% Score 55; DB 1; Length 245;
Best Local Similarity 53.8%; Pred. No. 6.08e+00;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 228 RIRYNNFTAPYTF 240
1 1:1:1111
0Y 3 RTLNNRYTGPYTF 15

RESULT 15
ID Y4ZC_RHISN STANDARD; PRT; 261 AA.
AC P55730;

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 28.3 KD PROTEIN Y4ZC.
GN Y4ZC.

OS Rhizobium sp. (strain NGR234).

OC Plasmid sym PNGR234a.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

RN 11)

RP SEQUENCE FROM N.A.

RX MEDLINE: 97305956.

RA Freiberg C.A., Fellay R., Batroch A., Broughton W.J., Rosenthal A.,

Perrut X.;

RT "Molecular basis of symbiosis between Rhizobium and legumes.";

CC -1- SIMILARITY: TO P. SYRINGAE (PV. PHASEOLICOLA) AVIRULENCE PROTEIN
CC AVIRPH3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000109; AAB91961.1; -.
KW Hypothetical protein; Plasmid.
SO SEQUENCE 261 AA: 28349 MW: 52939C1919DFEC43 CRC64;

Query Match 50.5% Score 55; DB 1; Length 261;
Best Local Similarity 60.0%; Pred. No. 6.08e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 238 RSLANRYSNP 247
1 1:1:111:1
0Y 3 RTLNNRYTGP 12

Search completed: Sat May 13 07:24:15 2000
Job time : 108 secs.

 WIRENET

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

21 56 51.4 129.3 042886 HYPOTHETICAL 14.9 KD P 1.27e+01
 22 56 51.4 212.14 002474 PUTATIVE RNA-DIRECTED 1.27e+01
 23 56 51.4 482.2 060051 ALPHA-AMYLASE PRECURSOR 1.27e+01
 24 56 51.4 648.12 080875 HYPOTHETICAL 74.0 KD P 1.27e+01
 25 56 51.4 771.14 080873 HYPOTHETICAL 87.3 KD P 1.27e+01
 26 56 51.4 843.11 092077 PUTATIVE TASTE RECEPTOR 1.27e+01
 27 56 51.4 1099.11 070248 MYOSIN IF 1.27e+01
 28 56 51.4 1823.5 026638 2 ALPHA FIBRILLAR COLL 1.27e+01
 29 56 51.4 2115.14 086500 M33 RNA FOR A NONSTRUC 1.27e+01
 30 56 51.4 345.10 09X194 F7A19.1 PROTEIN 1.92e+01
 31 56 51.4 380.10 082468 PROTEIN PHOSPHATASE-2C 1.92e+01
 32 56 51.4 468.3 074872 HYPOTHETICAL 52.6 KD P 1.92e+01
 33 56 51.4 550.5 09Y141 BCDNA-GH05741.1 1.92e+01
 34 56 51.4 2455.14 09WIC8 NONSTRUCTURAL POLIPROT 1.92e+01
 35 56 51.4 2501.14 09WIC9 NONSTRUCTURAL POLIPROT 1.92e+01
 36 56 51.4 49.5 120.11 061523 POLYMERASE POLYPROTEIN 2.89e+01
 37 56 51.4 49.5 229.14 09WIC6 REVERSE TRANSCRIPTASE 2.89e+01
 38 56 51.4 49.5 252.14 056251 29 KDA PROTEIN 2.89e+01
 39 56 51.4 49.5 405.1 026886 CONSERVED PROTEIN 2.89e+01
 40 56 51.4 49.5 707.11 061918 POL PROTEIN (FRAGMENT) 2.89e+01
 41 56 51.4 49.5 878.2 087985 HYPOTHETICAL 98.1 KD P 2.89e+01
 42 56 51.4 49.5 905.2 044488 NIFEN2 2.89e+01
 43 56 51.4 49.5 1180.14 09WHV7 POLYMERASE (FRAGMENT) 2.89e+01
 44 56 51.4 49.5 1204.14 09YK99 POL POLYPEPTIDE (FRAGM 2.89e+01
 45 56 51.4 49.5 1734.11 070355 GAG-POL POLYPROTEIN 2.89e+01

ALIGNMENTS

RESULT 1
 ID 035902 PRELIMINARY: PRT: 993 AA.

AC 035902: PRELIMINARY: PRT: 993 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE DESMOGLEIN 3 (FRAGMENT).

GN DSC3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-BALB/C;

RC ISHIKAWA H., LI K., UHTO J.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL: U86016; AAB5091.1; -

DR HSSP: P15116; INCI.

DR PROSITE: PS00232; CADHERIN: 2.

DR PFAM: PF00028; cadherin: 4.

FT Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.

FT NON-TER 993

SEQUENCE 993 AA: 107888 MW: 881794BD CRC32:

Query Match 61.5% Score 67: DB 11: Length 993;

Best Local Similarity 66.7% Pred No. 9.59e-02;

Matches 10: Conservative 2; Mismatches 2; Indels 1; Gaps 1;

DB 512 SVRTLDKRGKTYGPT 526

OY 1 SARTLN-NRYTGPYT 14

RESULT 2 PRELIMINARY: PRT: 1187 AA.

AC 093284: PRELIMINARY: PRT: 1187 AA.

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)

DE POL POLYPROTEIN (FRAGMENT).

GN POL.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

Statistics: Mean 26.282; Variance 35.962; scale 0.731

Database: splenb112
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mic 8:sp-oranalle
 9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Searched: 225878 seqs, 69334122 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Scoring table: PAM 150
 Gap 15

Sequence: 1 SARTLN-NRYTGPYT 15

Title: >US-08-991-628-6
 Description: (1-15) from US08991628.pep
 Perfect Score: 109

Tubular output not generated.

Run on: Sat May 13 07:24:34 2000; MasPar time 237.34 Seconds
 4.382 Million cell updates/sec

MSPrch_PP protein - protein database search, using Smith-Waterman algorithm

21 56 51.4 129.3 042886 HYPOTHETICAL 14.9 KD P 1.27e+01
 22 56 51.4 212.14 002474 PUTATIVE RNA-DIRECTED 1.27e+01
 23 56 51.4 482.2 060051 ALPHA-AMYLASE PRECURSOR 1.27e+01
 24 56 51.4 648.12 080875 HYPOTHETICAL 74.0 KD P 1.27e+01
 25 56 51.4 771.14 080873 HYPOTHETICAL 87.3 KD P 1.27e+01
 26 56 51.4 843.11 092077 PUTATIVE TASTE RECEPTOR 1.27e+01
 27 56 51.4 1099.11 070248 MYOSIN IF 1.27e+01
 28 56 51.4 1823.5 026638 2 ALPHA FIBRILLAR COLL 1.27e+01
 29 56 51.4 2115.14 086500 M33 RNA FOR A NONSTRUC 1.27e+01
 30 56 51.4 345.10 09X194 F7A19.1 PROTEIN 1.92e+01
 31 56 51.4 380.10 082468 PROTEIN PHOSPHATASE-2C 1.92e+01
 32 56 51.4 468.3 074872 HYPOTHETICAL 52.6 KD P 1.92e+01
 33 56 51.4 550.5 09Y141 BCDNA-GH05741.1 1.92e+01
 34 56 51.4 2455.14 09WIC8 NONSTRUCTURAL POLIPROT 1.92e+01
 35 56 51.4 2501.14 09WIC9 NONSTRUCTURAL POLIPROT 1.92e+01
 36 56 51.4 49.5 120.11 061523 POLYMERASE POLYPROTEIN 2.89e+01
 37 56 51.4 49.5 229.14 09WIC6 REVERSE TRANSCRIPTASE 2.89e+01
 38 56 51.4 49.5 252.14 056251 29 KDA PROTEIN 2.89e+01
 39 56 51.4 49.5 405.1 026886 CONSERVED PROTEIN 2.89e+01
 40 56 51.4 49.5 707.11 061918 POL PROTEIN (FRAGMENT) 2.89e+01
 41 56 51.4 49.5 878.2 087985 HYPOTHETICAL 98.1 KD P 2.89e+01
 42 56 51.4 49.5 905.2 044488 NIFEN2 2.89e+01
 43 56 51.4 49.5 1180.14 09WHV7 POLYMERASE (FRAGMENT) 2.89e+01
 44 56 51.4 49.5 1204.14 09YK99 POL POLYPEPTIDE (FRAGM 2.89e+01
 45 56 51.4 49.5 1734.11 070355 GAG-POL POLYPROTEIN 2.89e+01

RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J., COLE S.T.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
 RT lepreae";
 RT Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR EMBL: AL021287; CA16076.1; -.
 KW Hypothetical protein.
 SO SEQUENCE 163 AA; 18204 MW; D33CE4BB CRC32;

Query Match 54.1%; Score 59; DB 2; Length 163;
 Best Local Similarity 87.5%; Pred. No. 3.53e+00;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 111 ERYTGPYT 118
 :|||||
 OY 7 NRYTGPYT 14

RESULT 7
 ID 059397 PRELIMINARY; PRT; 330 AA.
 AC 059397;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE 330AA LONG HYPOTHEICAL N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE.
 GN PH1720.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE: 98344137.
 RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOGAWA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTUWA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KOSHIDA N., OGUCHI A.,
 RA AOI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
 RA KIKUCHI H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000007; BAA30834.1; -.
 DR PFAM: PF0118; Semialdehyde; 1.
 KW SEQUENCE 330 AA; 37168 MW; D80618ED CRC32;

Query Match 54.1%; Score 59; DB 1; Length 330;
 Best Local Similarity 53.8%; Pred. No. 3.53e+00;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 50 GGLNRLRTSDYNF 62
 :|:|:|:|:|
 OY 3 RLNNRYTGPYTF 15

RESULT 8
 ID 050378 PRELIMINARY; PRT; 360 AA.
 AC 050378;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1999 (TREMblrel. 12, Last annotation update)
 DE FERRIC EXOCHHELIN BIOSYNTHESIS.
 GN FXBA.
 OS Mycobacterium smegmatis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MC6;
 RX MEDLINE: 95191405.
 RA FISS E.H.;
 RT "Identification of genes involved in the sequestration of iron in
 RT mycobacteria: the ferric exochelin biosynthetic and uptake pathways.";
 RT Mol. Microbiol. 14:557-569(1994).
 DR EMBL: U10425; AAC43261.1; -.
 DR HSSP: P23882; 1FMD.
 DR PFAM: PF00551; formyl_transf. 1.
 KW SEQUENCE 360 AA; 41130 MW; A81D7F3B CRC32;

Query Match 54.1%; Score 59; DB 2; Length 360;
 Best Local Similarity 38.5%; Pred. No. 3.53e+00;
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 241 RALSDPYPNATY 253
 :|:|:|:|:|
 OY 3 RLNNRYTGPYTF 15

RESULT 9
 ID 066695 PRELIMINARY; PRT; 531 AA.
 AC 066695;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE HYPOTHEICAL 62.8 KD PROTEIN.
 GN AQ.372.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE: 98196666.
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBERK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus";
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBERK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE000686; AAC06652.1; -.
 DR KW Hypothetical protein.
 SO SEQUENCE 531 AA; 62830 MW; 585BE7E CRC32;

Query Match 54.1%; Score 59; DB 2; Length 531;
 Best Local Similarity 66.7%; Pred. No. 3.53e+00;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 80 TLKRLNLPYTF 91
 :|:|:|:|:|
 OY 4 TLNNRYTGPYTF 15

RESULT 10
 ID 001124 PRELIMINARY; PRT; 211 AA.
 AC 001124;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE COAT PROTEIN VP1 (FRAGMENT).
 GN ID.
 OS foot and mouth disease virus C.

OC VIRUSES: ssRNA positive-strand viruses, no DNA stage: Picornaviridae;
 OC Aphthovirus.
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 92260626.
 RA MARTINEZ M.A., DOPAZO J., HERNANDEZ J., MATEU M.G., SOBRINO F.,
 RA DOMINGO E., KNOWLES N.J.;
 RT "Evolution of the capsid protein genes of foot-and-mouth disease
 RT virus: antigenic variation without accumulation of amino acid
 RT substitutions over six decades.";
 RL J. Virol. 66:3357-3365(1992).
 CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -1- MISCELLANEOUS: THE COAT PROTEIN VP1 CONTAINS THE MAIN ANTIGENIC
 CC DETERMINANTS OF THE VIRION; THEREFORE, CHANGES IN ITS SEQUENCE
 CC MUST BE RESPONSIBLE FOR THE HIGH ANTIGENIC VARIABILITY OF THE
 CC VIRUS.
 DR EMBL: M90377: AAA91495.1: -.
 DR PFAM: PF00073: rhv: 1.
 KM COAT protein.
 FT NON_TER 1 1
 FT DOMAIN 134 157 ANTIGENIC A (BY SIMILARITY).
 FT DOMAIN 188 211 ANTIGENIC C (BY SIMILARITY).
 FT NON_TER 211 211
 SQ SEQUENCE 211 AA: 23028 MW: 818474E6 CRC32:
 Query Match 53.2%; Score 58; DB 14; Length 211;
 Best Local Similarity 53.8%; Pred. No. 5.43e+00;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 Db 125 RVLTATRYTGTTY 137
 Oy 3 RTLNRYTGPYTF 15
 RESULT 11 PRELIMINARY: PRT: 332 AA.
 ID 029749
 AC 029749;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TEMBLrel. 07, Last annotation update)
 DE NITRATE REDUCTASE, GAMMA SUBUNIT, PUTATIVE.
 GN AF0501.
 OS Archaeoglobus fulgidus.
 OC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae;
 OC Archaeoglobus.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLEIN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON C.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOTUS B.,
 RA PETERSON S., RITCH C.I., MCNEIL L.K., BADGER J.H., GLOER A., ZHOU L.,
 RA OVERBEER K., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., MOESE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL: AE001069: AAB90736.1: -.
 DR TIGR: AF0501: -.
 KW Hypothetical protein.
 SQ SEQUENCE 332 AA: 38386 MW: 6844D7B8 CRC32:
 Query Match 53.2%; Score 58; DB 1; Length 332;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 53 RTIYREDSPYT 64
 Oy 3 RTLNRYTGPYTF 14
 RESULT 12 PRELIMINARY: PRT: 506 AA.
 ID 017381
 AC 017381;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
 DE CE-FKH-1.
 GN CE-FKH-1.
 OS Caenorhabditis elegans.
 OC Eukaryota: Metazoa: Nematoda: Secernentea: Rhabditia: Rhabditidae;
 OC Rhabditina: Rhabditidae; Rhabditidae; Pelodertinae; Caenorhabditis.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-N2;
 RA AZARIA M., GOSACZYNSKI B., CHUNG M.A., KALB J.M., MCGHEE J.D.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MORTIMORE B.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U51163: AAA96319.1: -.
 DR EMBL: 292833: CAB07378.1: -.
 DR HSP: Q63245: ZHFH.
 DR PROSITE: PS00657: FORK_HEAD_1: 1.
 DR PROSITE: PS00658: FORK_HEAD_2: 1.
 DR PFAM: PF00250: Fork_head: 1.
 DR PRINTS: PR00053: FORKHEAD.
 SQ SEQUENCE 506 AA: 55426 MW: 54119AA3 CRC32:
 Query Match 53.2%; Score 58; DB 5; Length 506;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 443 TLNQVAPLY 454
 Oy 4 TLNNRYTGPYTF 15
 RESULT 13 PRELIMINARY: PRT: 770 AA.
 ID 007117
 AC 007117;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 87.2 KD PROTEIN.
 OS Barley stripe mosaic virus (BSMV).
 OC Viruses: ssRNA positive-strand viruses, no DNA stage: Hordeivirinae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9006400.
 RA KOZLOV YU.V., AFANAS'EV B.N., RUPASOV V.V., GOLOVA YU.B.,
 RA KULIAVA O.I., DOLYA V.V., ATABEKOV I.G., BAEV A.A.;
 RT "Primary structure of RNA 3 of barley stripe mosaic virus and its
 RT variability.";
 RL Mol. Biol. (Mosk) 23:1080-1090(1989).
 DR EMBL: X52774: CAA36983.1: -.
 DR PFAM: PF00978: RNA_dep_RNApol2: 1.
 KW Hypothetical protein.
 SQ SEQUENCE 770 AA: 87176 MW: 34477E66 CRC32:
 Query Match 53.2%; Score 58; DB 14; Length 770;
 Best Local Similarity 50.0%; Pred. No. 5.43e+00;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 720 KALVNRSPYS 731
 Oy 3 RTLNRYTGPYTF 14

RESULT 14
 ID 076025 PRELIMINARY: PRT: 80 AA.
 AC 076025;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94107601.
 RA MURPHY E., KORBER B.T., GEORGES-COURBOT M.C., YU B., PINTER A.,
 RA COOK D., KIENY M.P., GEORGES A., MATHIOT C., BARRE-SINOSSI F.,
 RA GIRARD M.;
 RT "Diversity of V3 region sequences of human immunodeficiency viruses
 RT type 1 from the central African Republic";
 RL AIDS Res. Hum. Retroviruses 9:997-1006(1993).
 DR EMBL: L11498; AAC37829.1; -.
 DR PFM: PF00516; GP120; 1.
 KM Envelope protein.
 FT NON TER 1 1
 SO SEQUENCE 80 AA: 9239 MM: 15E98296 CRC32;

Query Match 52.38; Score 57; DB 14; Length 80;
 Best Local Similarity 70.0%; Pred. No. 8.32e+00;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 63 NTRYTGSYTL 72
 | |||||
 Qy 6 NNRITGPTF 15

RESULT 15
 ID 005776 PRELIMINARY: PRT: 175 AA.
 AC 005776;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE SIMILAR TO MSF1 PROTEIN.
 GN L8167.12.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RX MEDLINE: 97313267.
 RA JOHNSTON M., HILLER L., RILES L., ALBERMANN K., ANDRE B., ANSORGE W.,
 RA BENES V., BRUCKNER M., DELIUS H., DUBOIS E., DUSTERHOFT A.,
 RA ENTIAN K.D., FLOETH M., GOFFEAU A., HEBLING U., HEUMANN K.,
 RA HEUS-NEITZEL D., HILBERT H., HILGER F., KLEINE K., KOTTER P.,
 RA LOUIS E.J., MESSENGUY F., MEMES H.W., MIOGA T., MOSTL D.,
 RA MULDER-AUER S., NENTWICH U., OBERMAIER B., PIRAVANDI E., POHL T.M.,
 RA PORTFELLE D., PURNELLE B., RECHMANN S., RIEGER M., RINKE M., ROSE M.,
 RA SCHARFE M., SCHERENS B., SCHOLLER P., SCHWAGER C., SCHWARZ S.,
 RA UNDERWOOD A.P., URRESTARAZU L.A., VANDENBOL M., VERHASSELT P.,
 RA VIERENDELS F., VOET M., VOLCKAERT G., VOSS H., WAMBUIT R., WEDLER E.,
 RA WEDLER H., ZIMMERMANN F.K., ZOLLNER A., HANI J., HOEISEL J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII";
 RL Nature 387:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA PAULEY A.;
 RN Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA WATERSTON R.;

RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA CHERRY J.M.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U14913; AAB67434.1; -.
 SO SEQUENCE 175 AA: 20108 MM: 9FCA0042 CRC32;

Query Match 52.38; Score 57; DB 3; Length 175;
 Best Local Similarity 46.2%; Pred. No. 8.32e+00;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 19 SRAFNRYPNPYS 31
 : : : : :
 Qy 2 ARTLNRYTGPYT 14

Search completed: Sat May 13 07:28:45 2000
 Job time : 251 secs.

THIS PAGE BLANK (USPTO)

 M E R E N
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 07:37:01 2000; Maspar time 3.01 Seconds
 118.060 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-991-628-7
 Description: (1-15) from US08991628.pep
 Perfect Score: 99
 Sequence: 1 OSGTMRTRHSTGCTN 15

Scoring table:
 PAM 150
 GAP 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35
 1:geneseqp

Statistics: Mean 17.183; Variance 54.742; scale 0.314

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	99	100.0	15	1	W04847	Self epitope of desmog	5.70e-04
2	99	100.0	999	1	R30742	Human pemphigus vulgar	6.17e-01
3	54	54.5	572	1	R06023	Viral haemagglutinin n	1.99e+02
4	49	49.5	140	1	R10423	Nod L gene product.	1.99e+02
5	49	49.5	392	1	Y03182	Topoisomerase II bindi	2.51e+02
6	48	48.5	802	1	R56550	Cold acclimatization p	2.51e+02
7	48	48.5	830	1	W80314	CITTA (class II transa	2.51e+02
8	48	48.5	987	1	W57057	Class II trans activat	2.51e+02
9	48	48.5	1085	1	P61048	A. nidulans phospheno	2.51e+02
10	48	48.5	1106	1	W80312	CITTA (class II transa	2.51e+02
11	48	48.5	1130	1	R72452	Class II transactivato	2.51e+02
12	48	48.5	1130	1	R81569	Class II transactivato	2.51e+02
13	48	48.5	1130	1	W57056	Class II trans activat	2.51e+02
14	48	48.5	1130	1	W80313	CITTA (class II transa	2.51e+02
15	48	48.5	1207	1	W80311	CITTA (class II transa	2.51e+02
16	47	47.5	483	1	W80214	Bacillus licheniformis	3.16e+02
17	47	47.5	483	1	R88449	Alpha-amylase H1331/N1	3.16e+02
18	47	47.5	483	1	R25344	Alpha-amylase mutant H	3.16e+02
19	47	47.5	483	1	R10579	Mutant alpha-amylase s	3.16e+02
20	47	47.5	483	1	R88450	Alpha-amylase H1331/T1	3.16e+02
21	47	47.5	483	1	R88451	Alpha-amylase V128E/H1	3.16e+02
22	47	47.5	483	1	R58734	M15V alpha-amylase.	3.16e+02
23	47	47.5	483	1	W05232	Bacillus licheniformis	3.16e+02

24	47	47.5	483	1	R58011	M197P alpha-amylase.	3.16e+02
25	47	47.5	483	1	R58729	M15H alpha-amylase.	3.16e+02
26	47	47.5	483	1	R57986	M197T alpha-amylase.	3.16e+02
27	47	47.5	483	1	R58737	M15E alpha-amylase.	3.16e+02
28	47	47.5	483	1	W57982	Alpha-amylase mutant M	3.16e+02
29	47	47.5	483	1	W57981	Alpha-amylase mutant M	3.16e+02
30	47	47.5	483	1	W73510	Alpha-amylase protein	3.16e+02
31	47	47.5	483	1	R58728	M15D alpha-amylase.	3.16e+02
32	47	47.5	483	1	R58730	M15K alpha-amylase.	3.16e+02
33	47	47.5	483	1	R58727	M15N alpha-amylase.	3.16e+02
34	47	47.5	483	1	R58725	M15A alpha-amylase.	3.16e+02
35	47	47.5	483	1	R58008	M197F alpha-amylase.	3.16e+02
36	47	47.5	483	1	R88442	Alpha-amylase N104D/H1	3.16e+02
37	47	47.5	483	1	W08204	B. licheniformis mutant	3.16e+02
38	47	47.5	487	1	W80215	M4 form alpha-amylase	3.16e+02
39	47	47.5	487	1	W05238	Bacillus licheniformis	3.16e+02
40	47	47.5	511	1	Y01583	An alpha-amylase (Am-L	3.16e+02
41	47	47.5	511	1	Y01582	An alpha-amylase prote	3.16e+02
42	47	47.5	512	1	Y00769	B. licheniformis alpha	3.16e+02
43	47	47.5	512	1	W73509	Alpha-amylase protein.	3.16e+02
44	47	47.5	512	1	W10325	B. licheniformis acid	3.16e+02
45	47	47.5	531	1	R98007	Pelb signal-Termamyl-1	3.16e+02

ALIGNMENTS

RESULT 1
 ID W04847; standard; peptide: 15 AA.
 AC W04847;
 DT 18-FEB-1997 (first entry)
 DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease.
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PE 07-MAR-1996; U03182.
 PR 07-MAR-1995; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Nucleic Acids Res 1991;19:101-10.
 DR WPI: 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1: Page 42: 58pp: English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 762-786)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04847-47.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 99; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5.70e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 OSGTMRTRHSTGCTN 15
 OY 1 OSGTMRTRHSTGCTN 15

RESULT 2
ID R30742 standard: Protein: 999 AA.
AC R30742:
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KM keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US798918.A.
PD 15-DEC-1992.
PR 27-NOV-1991: 798918.
PA (USSS) US DEPT HEALTH & HUMAN SERVICE.
PI Amagel M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
N-PSDB: Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure: Fig 7: 50pp: English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 99; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 5.70e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 762 QSGTMRTRHSTGGIN 776
QY 1 QSGTMRTRHSTGGIN 15
IIIIIIIIIIIIIIIIIIII

RESULT 3
ID R06023 standard: protein: 572 AA.
AC R06023:
DT 04-DEC-1990 (first entry)
DE Viral haemagglutinin neuraminidase.
KM Haemagglutinin neuraminidase: vaccine; parainfluenza; ds.
OS Bovine parainfluenza type III virus.
PN J02156863.A.
PD 15-JUN-1990.
PR 09-DEC-1988: 311655.
PA (JAPG) NIPPON ZEON KK.
DR WPI: 90-228484/30.
N-PSDB: Q05319.
PT Recombinant vaccinia virus - is which all or part of DNA coding
PT membrane fusion protein in combined to genom region.
PS Disclosure: p: Japanese.
CC Fragment of parainfluenza viral membrane fusion protein encoding DNA
CC may be incorporated into the vaccina virus, which may then be used
CC as a live vaccine for cows.
SQ Sequence 572 AA;

Query Match 54.5%; Score 54; DB 1; Length 572;
Best Local Similarity 53.3%; Pred. No. 6.17e+01;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 16 EPQTTRDRHSSKATN 30
QY 1 QSGTMRTRHSTGGIN 15
:::IIIIIIII::

RESULT 4
ID R10423 standard: Protein: 140 AA.
AC R10423:
DT 09-APR-1991 (first entry)
DE Nod L gene product.
KM Nod genes; parasite; toxin; promoter; nodulation; flavones; nod L.
OS Bradyrhizobium japonicum.

PN US4983519-A.
PD 08-JAN-1991.
PR 24-JUL-1987: 077561.
PA (STRAC/) STACEY G.
PI Stacey G, Schell MG, Nieuwkoop AJ, Deshmone NA, Banfalvi Z;
DR WPI: 91-036225/05.
N-PSDB: Q10325.
PT Nod genes and control sequences of Bradyrhizobium japonicum -
PT used for improving nodulation and for prodn. of foreign protein.
PS e.g. plant parasite toxin
PS Disclosure: Page 4: 11pp: English.
CC The nod L gene product may be used to restore nodulating ability
CC to Nod- mutants of Bradyrhizobium or Rhizobium strains and can
CC enhance nodulating ability.
SQ Sequence 140 AA;

Query Match 49.5%; Score 49; DB 1; Length 140;
Best Local Similarity 50.0%; Pred. No. 1.99e+02;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 85 SKRRRHATAGTK 96
QY 4 TMRTRHSTGGIN 15
:::IIIIIIII::

RESULT 5
ID Y03182 standard: protein: 392 AA.
AC Y03182:
DT 16-JUN-1999 (first entry)
DE Topoisomerase II binding protein 1.
KM Topoisomerase II binding protein; TopBP; anticancer agent.
OS Homo sapiens.
PN J11075856-A.
PD 23-MAR-1999.
PR 17-SEP-1997: 251544.
PA (CHUS) CHUGAI PHARM CO LTD.
PA (TSUR/) TSURUO T.
DR WPI: 99-257704/22.
N-PSDB: X28153.
PT New Topoisomerase II- binding protein - useful as an anticancer
PT agent
PS Claim 2: Page 13-14: 28pp: Japanese.
CC This sequence represents the topoisomerase II binding protein (TopBP) of
CC the invention. The TopBP protein is useful as an anticancer agent. TopBP
CC can be used as the target molecule for anticancer agent.
SQ Sequence 392 AA;

Query Match 49.5%; Score 49; DB 1; Length 392;
Best Local Similarity 66.7%; Pred. No. 1.99e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 1 MREKHDIGA 9
QY 5 MRRTRHSTG 13
IIIIIIII

RESULT 6
ID R56550 standard: Protein: 802 AA.
AC R56550:
DT 20-FEB-1995 (first entry)
DE Cold acclimatization protein CAP160 from spinach leaf.
KM Cold acclimatization; cold-tolerance; transgenic plant; drought-;
KM drought-resistance; drought-tolerance; crop improvement; CAP160;
KW freezing-tolerance.
OS Spinacia oleracea.
PN W09417186-A.
PD 04-AUG-1994.
PR 21-JAN-1994: U00581.
PA (UYFL) UNIV FLORIDA.
PI Guy CL, Haskell DW, Hofig A, Neven LG;

DR WPI: 94-264100/32.
 DR N-PSDB: Q70904.
 PT Cold acclimatization proteins CAP65 and CAP160 and their
 nucleotide sequences - used to confer increased cold tolerance
 PT and drought resistance on plants and microorganisms by genetic
 transformation
 PS Claim 6: Page 29: 49pp: English.
 CC Transgenic plants expressing this cold acclimatization
 CC protein have improved cold tolerance and/or drought-resistance.
 CC The cell expressing the protein is from the family Solanaceae,
 CC a citrus plant, a bacterium or a yeast cell.
 SO Sequence 802 AA.

Query Match 48.5%; Score 48; DB 1: Length 802;
 Best Local Similarity 53.3%; Pred. No. 2.51e+02;
 Matches 8: Conservative 1; Mismatches 6; Indels 0; Gaps 0;

DB 768 OPGTRRHIOGTE 782
 1 OSGTMRHSTGCTN 15

RESULT 7
 ID W80314 standard; Protein: 830 AA.
 AC W80314:
 DT 01-FEB-1999 (first entry)
 DE ClrA (class II transactivator) type IV protein.
 KM ClrA gene: class II transactivator; MHC class II molecule;
 KM Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 OS Homo sapiens.
 PN EP-874049-A1.
 PD 28-OCT-1998.
 PE 21-APR-1998: 400968
 PF 22-APR-1997: FR-004954.
 PA (TRGE) TRANSGENE SA.
 PR Mach B;
 DR WPI: 98-559115/48.
 PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
 PT e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Disclosure: Pages 66-69: 86pp: French.
 CC The present sequence represents type IV ClrA (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SO Sequence 830 AA.

Query Match 48.5%; Score 48; DB 1: Length 830;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 497 OPGTLRARQ 505
 1 OSGTMRH 9

RESULT 8
 ID W57057 standard; Protein: 987 AA.
 AC W57057:
 DT 17-AUG-1998 (first entry)
 DE Class II trans activator (ClrA) 151 deletion mutant.
 KM Class II trans activator; MHC; ClrA: autoimmune disease; treatment;
 KM transplantation; xenograft; major histocompatibility complex; diabetes;
 KM gene therapy; arthritis; ribozyme; 151 deletion mutant.
 OS Homo sapiens.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc_difference 6

FT /note- "wild type 143 amino acid residues deletion
 after this position"
 FT Misc_difference 7
 FT /label- E1501
 FT /note- "wild type Glu is replaced by Ile"
 FT Misc_difference 8
 FT /label- L1511
 FT /note- "wild type Leu is replaced by Ile"
 PN W09815626-A2.
 PD 16-APR-1998.
 PE 08-OCT-1997: G02751.
 PR 21-MAR-1997: GB-005911.
 PO 08-OCT-1995: GB-020940.
 PA (CHIL-) INST CHILD HEALTH.
 PI Fabre JW, Gustafsson KT, Yun S;
 DR WPI: 98-240813/21.
 DR N-PSDB: V28717.
 PT Deleted form of the class II trans-activator that reduces class II
 PT antigen expression - and ribozyme directed to trans-activator mRNA,
 PT related nucleic acid, vectors, transformed cells and antibodies,
 PT used for treating auto-immune disease and to inhibit xeno-graft
 PT rejection
 PS Claim 15: Pages 69-72: 104pp: English.
 CC This represents a 151 deletion mutant of the class II trans activator
 CC (ClrA) polypeptide. The N-terminal region of the wild type ClrA is
 CC deleted and can be used for reducing the expression of major
 CC histocompatibility complex (MHC) class II antigens. A ribozyme targeted
 CC to bases 1159-1161 (GUA) of human ClrA mRNA, or corresponding target in
 CC other species and the ClrA polypeptide can be used to reduce expression
 CC of MHC class II antigens, particularly for treatment of autoimmune
 CC disease (e.g. arthritis and diabetes) or to treat non-human animals
 CC intended as source of xenografts. The ribozyme and the ClrA polypeptide
 CC may be generated in vivo by gene therapy, using the encoding nucleic
 CC acids targeted for localised suppression of the immune response. Material
 CC from transgenic animals in which at least some cells are stably
 CC transfected with the ClrA encoding nucleic acids are used for animal-to-
 CC human transplantation. The ClrA polypeptide suppresses synthesis of
 CC class II antigens. In cells that express them constitutively or after
 CC lymphokine induction. It has no transcription activity but still binds to
 CC regulatory proteins in the promoter region of class II genes, so
 CC competitively inhibits endogenous ClrA.
 SO Sequence 987 AA.

Query Match 48.5%; Score 48; DB 1: Length 987;
 Best Local Similarity 55.6%; Pred. No. 2.51e+02;
 Matches 5: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 654 OPGTLRARQ 662
 1 OSGTMRH 9

RESULT 9
 ID P61048 standard; Protein: 1085 AA.
 AC P61048:
 DT 03-SEP-1991 (first entry)
 DE A.nidulans phenolol pyruvic acid carboxylase gene product.
 KM PPC: photosynthesis; pa181; pa171; pa172; pa172A.
 OS Anacystis nidulans.
 PN J61056082-A.
 PD 20-MAR-1986.
 PE 24-AUG-1984: 177352.
 PF 24-AUG-1984: JP-177352.
 PR (KOHZ/) KOHZUKI H.
 DR WPI: 86-115948/18.
 DR N-PSDB: N60838.
 PT PPC gene of cloned photosynthetic microorganisms - for
 PT intensifying photosynthesis carbon di-oxide-fixing ability of
 PT bacteria.
 PS Disclosure: Fig 1: 16pp: Japanese.
 CC Transforming a bacteria with the PPC encoding sequence intensifies
 CC CO2 fixation ability. Other PPC genes may also be found by
 CC hybridisation with an A.nidulans derived probe.

SQ Sequence 1085 AA

```
Query Match      48.5%; Score 48; DB 1; Length 1085;
Best Local Similarity 54.5%; Pred. No. 2.51e+02;
Matches 6; Conservative 2; Mismatches 3; Indels
```

Db	1054	TGLMRSRYSKG	10
		:	
QY	2	SGTMRTRHSTG	12

RESULT	10
ID	W80212
	CTD04774
	PROT012
	1105 AA

DT 01-FEB-1999 (first entry)
 DE C1rta (class II transactivator) type II protein.
 KW C1rta gene; class II transactivator; MHC class II molecule;
 KW Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 OS Homo sapiens.
 PN EP-874049-A1.
 PD 28-OCT-1998.
 PT 21-APR-1998; 400968.
 PR 22-APR-1997; FR-004954.
 RA (IRGE) TRANSGENE SA.
 PI Mach B.
 P1 WPI: 98-559115/48.
 PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
 PT e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Claim 10: Pages 58-61; 86pp; French.
 CC The present sequence represents type II C1rta (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines
 CC especially for cancer treatment.
 Q0 Sequence 1106 AA;

Query Match	48.5%	Score 48;	DB 1;	Length 1106;
Best Local Similarity	55.6%;	Pred. No. 2.51e+02;		
Matches	5;	Conservative	4;	Mismatches 0; Indels 0; Gaps 0

```
Db      773 QPGLRARQ 781
        1:1:1:1:
QY      1 QSGTMRTRH 9
```

RESULT	ID	Accession	Description
11	R72452	standard	Protein: 1130 AA.
AC	R72452		
DT	08-DEC-1995	(first entry)	
DE	Class II transactivator protein.		
KW	Class II transactivator protein; MHC class II;		
KW	Insulin dependent diabetes; multiple sclerosis; lupus erythematosus;		
KW	Rheumatoid arthritis; Kozak box; ATP/GTP binding cassette;		
KW	yeast RNA binding protein; transcription activation domain.		
OS	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	region	52..137	
FT	region	/note="acidic region rich in Glu and Asp"	
FT	region	163..195	
FT	region	/note="region I rich in Pro/Ser/Thr"	
FT	region	209..237	
FT	region	/note="region II rich in Pro/Ser/Thr"	
FT	region	261..322	
FT	region	/note="region III rich in Pro/Ser/Thr"	
FT	binding_site	420..426	
FT		/label="ATP/GTP binding site"	
EN	EP-648836-A.		
ED	19-APR-1995.		

PF 26-AUG-1994; 113378.
PR 26-AUG-1993; EP-113665

PA (MACH/) MACH B F.
PI Mach BF;
DR WPI; 95-148715/20.

PT Gene encoding a protein displaying class II trans-activator
PT activity - useful in compsn. for treating diseases associated
PT with impaired expression of MHC-II genes, etc.
PS Claim 11, Page 17-21; 32pp; English.
CC This protein sequence is encoded by the gene represented in 086110.
CC The protein is a class II transactivator (CIITA) essential for the
CC control of MHC class II gene expression in B lymphocytes. The DNA
CC sequence has two in frame start codons. The first of these is at position
CC 116 and acts as the translation initiation site. The second is present
CC (at position 188) in the context of a perfect Kozak box and may also
CC serve as a start codon for a protein of 1106 amino acids. The protein
CC contains three stretches rich in proline/serine/threonine. It also
CC contains a region rich in glutamate/aspartate (an acidic region) and an
CC ATP/GTP binding cassette. The acidic regions followed by three stretches
CC rich in proline, serine and threonine resembles a transcription
CC activation domain. There is also a leucine-rich region around amino
CC acids 979 to 1061 that shows weak homology with the N terminal portion of
CC a yeast RNA binding protein. The protein can be used for the
CC identification of inhibitors to proteins with CIITA activity. The
CC sequence can be used to treat diseases where a decrease in the level of
CC expression of MHC class II genes is desired e.g. insulin dependant
CC diabetes, multiple sclerosis, lupus erythematosus and rheumatoid
CC arthritis.
SO Sequence 1130 AA;

Query Match	48.5%	Score 48;	DB 1;	Length 1130;
Best Local Similarity	55.6%;	Pred. No. 2.51e+02;		
Matches	5;	Conservative	4;	Mismatches 0;
			Indels	0;
			Gaps	0

```
Db 797 QPGTLRARQ 805
QY 1 QSGTMRTRH 9
```

ID	RESULT	12
AC	R81569 standard; Protein; 1130 AA.	
AC	R81569:	
DE	22-MAY-1996 (first entry)	
DE	Class II transactivator.	
DE	CIITA: class II transactivator; transcription; MHC;	
KW	major histocompatibility complex class II; autoimmune disease;	
KW	therapy.	
OS	Homo sapiens.	
FT	key	Location/Qualifiers
FT	domain	26..352
FT	domain	/note="transcription activation domain"
FT	domain	301..1130
FT	domain	/note="interaction domain"
PN	MO9606107-A1.	
PD	29-FEB-1996.	
PR	22-AUG-1995; U10691.	
PR	24-AUG-1994; US-295502.	
FA	(HARD) HARVARD COLLEGE.	
PI	Douhan J, Glimcher LH, Zhou H;	
DR	WPI; 96-151325/15.	
DR	N-PSDB: 118028.	
PT	Methods which inhibit transcription activation by CIITA - causes	
PT	inhibition of MHC class II gene expression, used in therapy of	
PT	auto-immune disorders	
PS	Example 1; Page 36-41; 64pp; English.	
CC	A genomic DNA sequence (118028) codes for the class II	
CC	transactivator, CIITA (R81569), a protein essential for activation	
CC	of transcription of MHC class II genes. The CIITA transcription	
CC	activation domain provides useful information for identifying cpds.	
CC	which inhibit CIITA-dependent transcription. Such cpds are	
CC	potential autoimmune disease therapeutics by virtue of their ability	
CC	to inhibit transcription of the MHC class II genes. A second portion	

CC of CIITA, the interaction domain, mediates interaction of CIITA
CC and its target protein and activates MHC class II promoters.
CC Cpd's that inhibit this binding are also potential autoimmune
CC disease therapeutics.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

RESULT 13
ID W57056 standard; Protein: 1130 AA.
AC W57056;
DT 17-AUG-1998 (first entry)
DE Class II trans activator (CIITA) polypeptide.
KM Class II trans activator; MHC; CIITA; autoimmune disease; treatment;
KW transplantation; xenograft; major histocompatibility complex; diabetes;
KW gene therapy; arthritis; ribozyme.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD WO9815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

QY 1 QSGTMRTRH 9

RESULT 14
ID W80313 standard; Protein: 1130 AA.
AC W80313;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type III protein.
KM CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD WO9815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 1; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 797 QPGTLRARQ 805
1 QSGTMRTRH 9

RESULT 15
ID W80311 standard; Protein: 1207 AA.
AC W80311;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type I protein.
KM CIITA gene; class II transactivator; MHC class II molecule;
KW interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
FH Key
FT Region 52..137 Location/Qualifiers
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PD WO9815626-A2.
PN 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JW, Gustafsson KT, Yun S;
DR WPI: 98-240813/21.
DR N-PSDB: V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens, in cells that express them
CC constitutively or after lymphokine induction. It has no transcription
CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SO Sequence 1207 AA;

Query Match 48.5%; Score 48; DB 1; Length 1207;
Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 55.6%; Pred. No. 2.51e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 874 QPGLRARQ 882
I:|:|:|:|:

Oy 1 QSGTMRTRH 9

Search completed: Sat May 13 07:37:08 2000
Job time : 7 secs.

Db 762 OSGTMRHSTGCTN 776
|||||
Oy 1 OSGTMRHSTGCTN 15

RESULT 2
ENTRY BVFSL #type complete
TITLE SOL protein, large splice form - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999

ACCESSIONS
REFERENCE A41146
#authors Delaney, S.J.; Hayward, D.C.; Barleben, F.; Fischbach, K.F.; Miklos, G.L.G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7214-7218
#title Molecular cloning and analysis of small optic lobes, a structural brain gene of Drosophila melanogaster.
#cross-references MUID:91334436
#accession A41146
#molecule_type mRNA
#residues 1-1597 #label DEL

COMMENT #cross-references GB:M64084; NID:92760822; PID:AA95431.1; PID:9158483
The sol (small optic lobes) mutation eliminates certain classes of columnar neurons.
COMMENT An alternate splice form of 395 amino acids is observed, in which the first 393 are identical to the large sol protein.

GENETICS
#gene sol
#cross-references FlyBase:FBgn0003464
#map_position 19F4
#classification #superfamily sol protein; calpain catalytic domain homology
KEYWORDS #alternative splicing; brain; zinc finger
FEATURE
12-29 #region zinc finger CCCC motif
141-158 #region zinc finger CCCC motif
649-667 #region zinc finger CCCC motif
673-702 #region glutamine/histidine-rich
713-730 #region zinc finger CCCC motif
754-771 #region zinc finger CCCC motif
936-953 #region zinc finger CCCC motif
1047-1307 #domain calpain catalytic domain homology #label CALP
1082,1248,1268 #active_site Cys, His, Asn #status predicted
SUMMARY #length 1597 #molecular_weight 174713 #checksum 8253

Query Match 57.6%; Score 57; DB 1; Length 1597;
Best Local Similarity 61.5%; Pred. No. 3,67e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 873 SGAIPKRRHSTGGS 885
|||||
Oy 2 SGAIPKRRHSTGCT 14

RESULT 3
ENTRY T08428 #type complete
TITLE gene small optic lobes protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999

ACCESSIONS
REFERENCE T08428
#authors Maleczka, R.; de Couet, H.G.; Miklos, G.L.
#journal Proc. Natl. Acad. Sci. U.S.A. (1998) 95:3731-3736
#title Data transferability from model organisms to human beings: insights from the functional genomics of the flightless region of Drosophila.
#cross-references MUID:98188272
#accession T08428
#status translated from GB/EMBL/DBJ

Db 873 SGAIPKRRHSTGGS 885
|||||
Oy 2 SGAIPKRRHSTGCT 14

RESULT 4
ENTRY S33886 #type fragment
TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain
ORGANISM #formal_name Euplotes octocarinatus
DATE 08-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 12-Sep-1997

ACCESSIONS
REFERENCE S33886; S70414; S73103
#authors Kaufmann, J.; Klein, A.
#journal Nucleic Acids Res. (1992) 20:4445-4450
#title Gene dosage as a possible major determinant for equal expression levels of genes encoding RNA polymerase subunits in the hypotrichous ciliate Euplotes octocarinatus.
#cross-references MUID:93027138
#accession S33886
#status translation not shown

GENETICS
#gene RPB1
#genetic_code SGC9
#introns 30/3; 74/1
CLASSIFICATION #superfamily human DNA-directed RNA polymerase II largest chain
KEYWORDS DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc finger
SUMMARY #length 478 #checksum 3090

Query Match 56.6%; Score 56; DB 2; Length 478;
Best Local Similarity 53.8%; Pred. No. 6,02e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 301 QAGQVNRHRSKSG 313
|||||
Oy 1 QSGTMRHSTGCT 13

RESULT 5
ENTRY A44991 #type complete

```

TITLE      protein-tyrosine kinase (EC 2.7.1.112)  fyn - mouse
ALTERNATE_NAMES
ORGANISM   Mus musculus #common_name house mouse
DATE       03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
18-Jun-1999

ACCESSIONS
REFERENCE  A44991
           A44991
           Cooke, M.P.; Perlmutter, R.M.
           New Biol. (1989) 1:66-74
           Expression of a novel form of the fyn proto-oncogene in
           hematopoietic cells.
           #cross-references MUID:91175680
           #accession A44991
           #status preliminary
           #molecule_type mRNA
           #residues 1-534 #label COO
           #cross-references GB:M27266; NID:9193357; PIDN:AAA37644.1; PID:9309241
           In the authors' translation an additional residue Leu
           was shown after Trs, for residue 459

CLASSIFICATION
KEYWORDS   #superfamily protein-tyrosine kinase src; protein kinase
           #homology: SH2 homology: SH3 homology
           ATP; autophosphorylation; phosphoprotein; phosphotransferase;
           tyrosine-specific protein kinase

FEATURE
FEATURE    89-138      #domain SH3 homology #label SH3\
           149-246      #domain SH2 homology #label SH2\
           266-524      #domain protein kinase homology #label KIN\
           274-282      #region protein kinase ATP-binding motif
SUMMARY    #length 534 #molecular_weight 60057 #checksum 2574

Query Match      55.66; Score 55; DB 2; Length 534;
Best Local Similarity 57.18; Pred. No. 9,93e-01;
Matches      8; Conservative      4; Mismatches      2; Indels      0; Gaps      0;

Db      72 HTGTLRTGCGTGT 85
Oy      1 QSGTMRTRHSTGCT 14

RESULT      6
ENTRY
TITLE      S33568      #type complete
ORGANISM   protein-tyrosine kinase (EC 2.7.1.112)  fyn - chicken
DATE       31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change
18-Jun-1999
ACCESSIONS S33568; S36351; S20807
REFERENCE  S33568
           Sudol, M.; Graulich, H.; Newman, L.; Sarkar, A.; Sukegawa,
           J.; Yamamoto, T.
           Oncogene (1993) 8:823-831
           A novel ves-related kinase, Yrk, is expressed at elevated
           levels in neural and hematopoietic tissues.
           #cross-references MUID:93205395
           #accession S33568
           #molecule_type mRNA
           #residues 1-534 #label SUD1
           #cross-references EMBL:X52841
REFERENCE  S36351
           Sudol, M.
           submitted to the EMBL Data Library, May 1990
           #accession S36351
           #molecule_type mRNA
           #residues 1-348; '1', 350-534 #label SUD2
           #cross-references EMBL:X52841; NID:962861; PIDN:CAA37025.1; PID:962862

GENETICS
CLASSIFICATION
KEYWORDS   #superfamily protein-tyrosine kinase src; protein kinase
           #homology: SH2 homology: SH3 homology
           ATP; phosphoprotein; phosphotransferase; proto-oncogene;
           transforming protein; tyrosine-specific protein kinase

FEATURE
FEATURE    89-138      #domain SH3 homology #label SH3\
           149-246      #domain SH2 homology #label SH2\

```

```

266-524      #domain protein kinase homology #label K1V\
274-282      #region protein kinase ATP-binding motif\
296          #binding_site ATP (lys) #status predicted\
477,528      #binding_site phosphate (tyr) (covalent) #status
              predicted
SUMMARY      #length 534 #molecular-weight 60267 #checksum 4864

Query Match    57.6% Score 55: DB 2: Length 534;
Best Local Similarity 57.1%; Pred. No. 9,93e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db       72 HTGTLTRGTGCT 85
Oy       1 OSGTMRTRHSTGCT 14
        ::::|::|::|::|
RESULT   7 TVHUSY      #type complete
ENTRY    protein-tyrosine kinase (EC 2.7.1.112) fyn, splice form B -
TITLE     human
ALTERNATE_NAMES
ORGANISM  fyn; p59(fyn); protein-tyrosine kinase slk; protein-tyrosine
           kinase syn; proto-oncogene fyn
           #formal_name Homo sapiens #common_name Man
DATE      31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
           11-Jun-1999
ACCESSION A243314; A25389; I59120; I79512
REFERENCE A243314
#authors Semba, K.; Nishizawa, M.; Miyajima, N.; Yoshida, M.C.;
           Sukegawa, J.; Yamashita, Y.; Sasaki, M.; Yamamoto, T.;
           Toyoshima, K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:5459-5463
#title yes-related protooncogene, syn, belongs to the
           protein-tyrosine kinase family.
#cross-references GB:M14333; NID:g181171; PIDN:AAC08285.1; PID:g181172
#accession A243314
#molecule_type mRNA
#residues 1-537 ##label SEM
REFERENCE #cross-references GB:M14333; NID:g181171; PIDN:AAC08285.1; PID:g181172
A25389
#authors Kawakami, T.; Pennington, C.Y.; Robbins, K.C.
#journal Mol. Cell. Biol. (1986) 6:4195-4201
#title Isolation and oncogenic potential of a novel human src-like
           gene.
#cross-references MUID:87089775
#accession A25389
#molecule_type mRNA
#residues 1-183,'S',185-436,'R',438-537 ##label KAW1
#note #cross-references GB:M14676; NID:g338227; PIDN:AAA36615.1; PID:g338228
the authors translated the codon GAA for residue 265 as
Gln and GGA for residue 278 as Gln
I59120
REFERENCE #cross-references GB:M20284; NID:g182842; PIDN:AAA52492.1; PID:g182843
Kawakami, T.; Kawakami, Y.; Aaronson, S.A.; Robbins, K.C.
#journal Proc. Natl. Acad. Sci. U.S.A. (1988) 85:3870-3874
#title Acquisition of transforming properties by FYN, a normal
           SRC-related human gene.
#cross-references MUID:88234523
#accession I59120
#status translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 514-522 ##label KAW2
#cross-references GB:M20284; NID:g182842; PIDN:AAA52492.1; PID:g182843
#experimental_source clone pfyn(c)-11; termination mutant p56(fyn)
#accession I79512
#status translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 514-524,'ORS' ##label KAW3
#cross-references GB:M20285; NID:g182844; PIDN:AAA52492.1; PID:g182845
#experimental_source clone pfyn(c)-21; recombinant mutant p57(fyn)
AS8826
REFERENCE #cross-references GB:M20285; NID:g182844; PIDN:AAA52492.1; PID:g182845
A58826
#authors Peters, D.J.; McGrew, B.R.; Petron, D.C.; Lipatak, L.M.;
           Landano, A.P.
#journal Oncogene (1990) 5:1313-1319
#title In vivo phosphorylation and membrane association of the fyn

```

proto-oncogene product in IM-9 human lymphoblasts.
#cross-references MUID:91016431
#contents annotation: myristylation: Tyr-531 phosphorylation
REFERENCE A51398
#authors Noble, M.; Musacchio, A.; Saraste, M.; Wierenga, R.
#submission submitted to the Brookhaven Protein Data Bank, May 1993
#cross-references PDB:1SHF
#contents annotation: X-ray crystallography, 1.9 angstroms, residues 84-142

REFERENCE A58827
#authors Noble, M.E.M.; Musacchio, A.; Saraste, M.; Courtneidge, S.A.; Wierenga, R.K.
#journal EMBO J. (1993) 12:2617-2624
#title Crystal structure of the SH3 domain in human Fyn: comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin.
#cross-references MUID:93327750
#contents annotation: X-ray crystallography, 1.9 angstroms
REFERENCE A65653
#authors Musacchio, A.; Saraste, M.; Wilmanns, M.
#submission submitted to the Brookhaven Protein Data Bank, May 1995
#cross-references PDB:1FTN
#contents annotation: X-ray crystallography, 2.3 angstroms, residues 81-142

REFERENCE A58828
#authors Musacchio, A.; Saraste, M.; Wilmanns, M.
#journal Nature Struct. Biol. (1994) 1:546-551
#title High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides.
#cross-references MUID:95393198
#contents annotation: X-ray crystallography, 2.3 angstroms
REFERENCE A66268
#authors Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYF
#contents annotation: conformation by (1)H-NMR, residues 84-141

REFERENCE A66269
#authors Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYG
#contents annotation: conformation by (1)H-NMR, residues 84-141

GENETICS
#gene GDB:FYN
#cross-references GDB:118797; OMIM:137025
#map_position 6q21-6q21
COMPLEX monomer
FUNCTION
#description catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
CLASSIFICATION
#superfamily protein-tyrosine kinase src: protein kinase
#homology: SH2 homology; SH3 homology
#formal_name: SH2 homology; SH3 homology
KEYWORDS
#alternative splicing: ATP; autophosphorylation; blocked amino end; lipoprotein; monomer; myristylation; phosphoprotein; phosphotransferase; proto-oncogene; thioester bond; transforming protein; tyrosine-specific protein kinase

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif\
#modified_site myristylated amino end (Gly) (1n mature form) #status predicted\
#binding_site palmitate (Cys) (covalent) #status predicted\
#active_site Lys #status predicted\
#binding_site phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted\
#binding_site phosphate (Tyr) (covalent) #status experimental

SUMMARY
#length 537 #molecular-weight 60762 #checksum 9747
Query Match 55.6%; Score 55; DB 1; Length 537;
Best Local Similarity 57.1%; Pred. NO. 9.93e-01;

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTLRTRGCGT 85
:::|||||::|
Qy 1 QSGTMRTRHSTGCT 14

RESULT 8
ENTRY I51592 #type complete
TITLE p59(Xfyn) - Xiphophorus helleri
#formal_name Xiphophorus helleri
ORGANISM
DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 18-Jun-1999

ACCESSIONS I51592
REFERENCE I51592
#authors Hannig, G.; Ottile, S.; Scharf, M.
#journal Oncogene (1991) 6:361-369
#title Conservation of structure and expression of the c-yes and fyn genes in lower vertebrates.
#cross-references MUID:9187435
#accession I51592
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-537 #label HAN
#cross-references EMBL:X54971; NID:g64481; PIDN:CAA38715.1; PID:g64482

GENETICS
#gene Xfyn
#cross-references EMBL:X54971; NID:g64481; PIDN:CAA38715.1; PID:g64482

CLASSIFICATION
#superfamily protein-tyrosine kinase src: protein kinase
#homology: SH2 homology; SH3 homology

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN
SUMMARY
#length 537 #molecular-weight 60447 #checksum 621
Query Match 55.6%; Score 55; DB 2; Length 537;
Best Local Similarity 57.1%; Pred. NO. 9.93e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTLRTRGCGT 85
:::|||||::|
Qy 1 QSGTMRTRHSTGCT 14

RESULT 9
ENTRY A43806 #type complete
TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - African clawed frog
ORGANISM
DATE 30-Jan-1993 #sequence_revision 30-Jan-1993 #text_change 18-Jun-1999

ACCESSIONS A43806
REFERENCE A43806
#authors Steele, R.E.; Deng, J.C.; Ghosh, C.R.; Fero, J.B.
#journal Oncogene (1990) 5:369-376
#title Structure and expression of fyn genes in Xenopus laevis.
#cross-references MUID:90191723
#accession A43806
#status Preliminary
#molecule_type mRNA
#residues 1-537 #label STE
#cross-references GB:X52188; GB:X52189; NID:g644701; PIDN:CAA36435.1; PID:g644702

CLASSIFICATION
#superfamily protein-tyrosine kinase src: protein kinase
#homology: SH2 homology; SH3 homology
#formal_name SH2 homology; SH3 homology
KEYWORDS
#ATP; autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE
89-138 #domain SH3 homology #label SH3\
149-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif
SUMMARY
#length 537 #molecular-weight 60846 #checksum 2563

Query Match 55.6%; Score 55; DB 2; Length 537;
 Best Local Similarity 57.1%; Pred. No. 9.93e+01;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTRTGTGTGT 85
 1 OSGMTTRHSTGCT 14

RESULT 10
 ENTRY 70369 #type complete
 TITLE gene m3 protein - Lactococcus phage b1170
 ORGANISM #formal_name Lactococcus phage b1170
 DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS 70369
 REFERENCE 214903
 #authors Cruz-Le Cog, A.M.; Cesselin, B.; Commissaire, J.; Anba, J.; Kyriakidis, S.; Chopin, M.C.
 #submission Submitted to the EMBL Data Library, June 1997
 #description Sequence and organization of the lactococcal isometric b1170 phage genome.
 #accession T03369
 #status preliminary; translated from GB/EMBL/DBJ
 #molecule_type DNA
 #residues 1-160 #label CRU
 #cross-references EMBL:AF009630; NID:g3282260; PID:g3282321

GENETICS m3
 SUMMARY #length 160 #molecular-weight 18072 #checksum 3777

Query Match 53.5%; Score 53; DB 2; Length 160;
 Best Local Similarity 60.0%; Pred. No. 2.63e+00;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 90 GVMRRHNIG 99
 3 GTMTRHSTG 12

RESULT 11
 ENTRY 713827 #type complete
 TITLE kinesin-73 - fruit fly (Drosophila melanogaster)
 ORGANISM #formal_name Drosophila melanogaster
 DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

ACCESSIONS 713827
 REFERENCE 217784
 #authors Li, H.P.; Liu, Z.M.; Nirenberg, M.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1997) 94:1086-1091
 #title Kinesin-73 in the nervous system of Drosophila embryos.
 #accession 713827
 #status preliminary; translated from GB/EMBL/DBJ
 #molecule_type mRNA
 #residues 1-1921 #label LTH
 #cross-references EMBL:U01788; NID:g1906595; PID:g1906596; PIDN:AA0504.1

SUMMARY #length 1921 #molecular-weight 215047 #checksum 9262

Query Match 52.5%; Score 52; DB 2; Length 1921;
 Best Local Similarity 58.3%; Pred. No. 4.24e+00;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1431 GGVRSRHSFGG 1442
 2 SGTMRHSTG 13

RESULT 12
 ENTRY JN0578 #type complete
 TITLE p-Aminobenzoic acid synthetase (EC 4.1.3.-) B - Streptomyces lividans

ORGANISM #formal_name Streptomyces lividans
 DATE 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 22-Jun-1999

ACCESSIONS JN0578
 REFERENCE JN0577
 #authors Arhin, F.F.; Vining, L.C.
 #journal Gene (1993) 126:129-133.
 #title Organization of the genes encoding p-aminobenzoic acid synthetase from Streptomyces lividans 1326.
 #cross-references MUID:93231526
 #accession JN0578
 #molecule_type DNA
 #residues 1-475 #label ARH
 #cross-references GB:M64059; NID:g153393; PIDN:AAA26796.1; PID:g153394
 COMMENT This enzyme participates in p-aminobenzoic acid biosynthesis.

GENETICS
 #gene PabB
 CLASSIFICATION #superfamily anthranilate synthase component I
 KEYWORDS carbon-carbon lyase; oxo-acid-lyase; p-aminobenzoate biosynthesis

SUMMARY #length 475 #molecular-weight 52042 #checksum 6682

Query Match 51.5%; Score 51; DB 2; Length 475;
 Best Local Similarity 42.9%; Pred. No. 6.79e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 281 ACTLGRPRAGCD 294
 2 SGTMRHSTG 15

RESULT 13
 ENTRY S04605 #type complete
 TITLE glycinin G3 - soybean
 ORGANISM #formal_name Glycine max #common_name soybean
 DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Aug-1999

ACCESSIONS S04605
 REFERENCE S04605
 #authors Cho, T.J.; Nielsen, N.C.
 #journal Nucleic Acids Res. (1989) 17:4388
 #title The glycinin G3 gene from soybean.
 #cross-references MUID:89296500
 #accession S04605
 #molecule_type DNA
 #residues 1-481 #label CHO
 #cross-references EMBL:X15123; NID:g18638; PIDN:CAA33217.1; PID:g18639
 #experimental_source variety Date

GENETICS
 #gene GY3
 #introns 96/1: 177/3: 352/3
 CLASSIFICATION #superfamily glycinin
 SUMMARY #length 481 #molecular-weight 54241 #checksum 2269

Query Match 51.5%; Score 51; DB 2; Length 481;
 Best Local Similarity 58.3%; Pred. No. 6.79e+00;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 304 TMRHSTG 315
 4 TMRHSTG 15

RESULT 14
 ENTRY S11003 #type complete
 TITLE glycinin G3 precursor - soybean
 ORGANISM #formal_name Glycine max #common_name soybean
 DATE 21-Nov-1993 #sequence_revision 19-Jan-1996 #text_change 03-May-1996

ACCESSIONS S11003
 REFERENCE S10851
 #authors Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.; Scallan, B.J.; Fischer, R.L.; Sims, T.L.; Drews, G.N.;

Goldberg, R.B.
#journal Plant Cell (1989) 1:313-328
#title Characterization of the glycylcin gene family in soybean.
#cross-references MUID:92393391
#accession S11003
#molecule_type DNA
#residues 1-484 #label NIE
##experimental_source variety Dare

GENETICS
#gene Gy3
CLASSIFICATION #superfamily glycylcin
KEYWORDS storage protein
FEATURE
1-19
20-484 #domain signal sequence #status predicted #label SIG\
SUMMARY #product glycylcin G3 #status predicted #label MAT\
#length 484 #molecular-weight 54359 #checksum 8616

Query Match 51.5%; Score 51; DB 2; Length 484;
Best Local Similarity 58.3%; Pred. No. 6.79e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 304 TMRLRHNGOTS 315
||| ||: ||:
QY 4 TMRTRHSTGCTN 15

RESULT 15
ENTRY FMSYC2 #type complete
TITLE glycylcin chain A1abx precursor - soybean
ALTERNATE_NAMES 115 globulin; glycylcin A1ab1b
ORGANISM #formal_name glycine max #common_name soybean
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
A23497; S10502

ACCESSIONS
REFERENCE A23497
#authors Negoro, T.; Momma, T.; Fukazawa, C.
#journal Nucleic Acids Res. (1985) 13:6719-6731
#title A cDNA clone encoding a glycylcin A-1a subunit precursor of soybean.
#cross-references MUID:86041867
#accession A23497
#molecule_type mRNA
#residues 1-495 #label NEG
##cross-references GB:X02985; NID:g18614; PIDN:CAA26723.1; PID:g18615
##experimental_source cv. Bonminor1
##note the authors translated the codon AAC for residue 449 as lys
##note because of current nomenclature ambiguities, the authors choose to designate the basic chain as Bx

REFERENCE S10502
#authors Kitamura, Y.; Arahira, M.; Itoh, Y.; Fukazawa, C.
#journal Nucleic Acids Res. (1990) 18:4245
#title The complete nucleotide sequence of soybean glycylcin A2B1a gene spanning to another glycylcin gene A1ab1b.
#cross-references MUID:90332420
#accession S10502
#status preliminary: translation not shown
#molecule_type DNA
#residues 481-495 #label KIT
##cross-references EMBL:X53404; NID:g18522; PIDN:CAA37479.1; PID:g18523
COMMENT The source of this protein was cotyledon tissue taken 38 days after flowering.
COMMENT By homology with the A2B1a component, residues 307-310, and 491-495 are removed from the precursor during posttranslational processing.

CLASSIFICATION #superfamily glycylcin
KEYWORDS seed; storage protein
FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-306 #product glycylcin chain A1a #status predicted #label GLA\
311-490 #product glycylcin chain Bx #status predicted #label GLB\
107-317 #disulfide_bonds #status predicted

SUMMARY #length 495 #molecular-weight 55506 #checksum 8518
Query Match 51.5%; Score 51; DB 1; Length 495;
Best Local Similarity 58.3%; Pred. No. 6.79e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 318 TMRLRHNGOTS 329
||| ||: ||:
QY 4 TMRTRHSTGCTN 15

Search completed: Sat May 13 07:36:42 2000
Job time : 8 secs.

 [W][O][R][L][D] (TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 07:29:58 2000; Maspar time 89.82 seconds

Tabular output not generated. 5.086 Million cell updates/sec

Title: >US-08-991-628-7

Description: (1-15) from US08991628.pep

Sequence: 1 OSCGTRMRHSTGCTN 15

Scoring table: PAM 150

Gap 15

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 24.519; Variance 27.183; scale 0.902

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	99	100.0	999	1	DSG3_HUMAN	DESMOGLEIN 3 PRECURSOR	3.88e-13
2	57	57.6	1597	1	SOL_DROME	SMALL OPTIC LOBES PROT	1.21e-01
3	56	56.6	478	1	RPI1_EUPOC	DNA-DIRECTED RNA POLYM	2.08e-01
4	55	55.6	533	1	FYN_MOUSE	PROTO-ONCOGENE TYROSIN	3.55e-01
5	55	55.6	533	1	FYN_MOUSE	PROTO-ONCOGENE TYROSIN	3.55e-01
6	55	55.6	533	1	FYN_MOUSE	PROTO-ONCOGENE TYROSIN	3.55e-01
7	55	55.6	533	1	FYN_MOUSE	PROTO-ONCOGENE TYROSIN	3.55e-01
8	55	55.6	533	1	FYN_MOUSE	PROTO-ONCOGENE TYROSIN	3.55e-01
9	52	52.5	543	1	TCPH_HUMAN	T-COMPLEX PROTEIN 1, E	1.69e+00
10	51	51.5	475	1	PAB3_HUMAN	PARA-AMINOBENZOCARBYN	2.80e+00
11	51	51.5	481	1	GLC1_SOYBN	GLYCININ G1 PRECURSOR	2.80e+00
12	51	51.5	495	1	GLC1_SOYBN	GLYCININ G1 PRECURSOR	2.80e+00
13	50	50.5	120	1	MASD_SCHUA	IMMUNOGENIC MIRACIDIAL	4.61e+00
14	50	50.5	286	1	AAC8_STRFR	AMINOGLYCOSIDE N3'-ACE	4.61e+00
15	50	50.5	572	1	HEMA_P13B	HEMAGGLUTININ-NEURAMIN	4.61e+00
16	50	50.5	2318	1	NTC3_MOUSE	NEUROGENIC LOCUS NOTCH	4.61e+00
17	49	49.5	230	1	RL4_MYCLE	50S RIBOSOMAL PROTEIN	7.54e+00
18	48	48.5	172	1	SCUD_MAGGR	SCYTALONE DEHYDRATASE	1.22e+01
19	48	48.5	291	1	T2EB_HUMAN	TRANSCRIPTION INITIATI	1.22e+01
20	48	48.5	355	1	COAT_AEDER	COAT PROTEIN VP1 [CONT	1.22e+01
21	48	48.5	462	1	BIND_LYTVB	BINDIN PRECURSOR.	1.22e+01
22	48	48.5	471	1	NRAM_IAPFW	NEURAMINIDASE (EC 3.2.	1.22e+01
23	48	48.5	510	1	KR2_VAVD	PROBABLE SERINE/THREON	1.22e+01

24	48	48.5	510	1	CATB_PSESY	CATALASE PRECURSOR (EC	1.22e+01
25	48	48.5	729	1	NARB_SYN7	NITRATE REDUCTASE (EC	1.22e+01
26	48	48.5	823	1	UBPG_HUMAN	UBIQUITIN CARBOXYL-TER	1.22e+01
27	48	48.5	927	1	CC15_SCHPO	CELL DIVISION CONTROL	1.22e+01
28	48	48.5	1053	1	CAPP_ANANI	PHOSPHONOLPYRUVATE CA	1.22e+01
29	48	48.5	1125	1	PHYA_POPTM	PHYTOCHROME A.	1.22e+01
30	48	48.5	1130	1	C2TA_HUMAN	MHC CLASS II TRANSACTI	1.22e+01
31	48	48.5	1201	1	COPA_YEAST	CONTOMER ALPHA SUBUNIT	1.22e+01
32	48	48.5	1591	1	TIM_MOUSE	T-LYMPHOMA INVASION AN	1.22e+01
33	48	48.5	1591	1	TIM_MOUSE	T-LYMPHOMA INVASION AN	1.22e+01
34	48	48.5	220	1	R8B_ORYSA	40S RIBOSOMAL PROTEIN	1.97e+01
35	47	47.5	221	1	RS8_MAIZE	40S RIBOSOMAL PROTEIN	1.97e+01
36	47	47.5	370	1	MAF2_MOUSE	TRANSCRIPTION FACTOR M	1.97e+01
37	47	47.5	416	1	FLIA_HUMAN	FLIAGRIN PRECURSOR (F	1.97e+01
38	47	47.5	469	1	NRAM_IAMCO	NEURAMINIDASE (EC 3.2.	1.97e+01
39	47	47.5	469	1	NIFN_RHISN	NITROGENASE IRON-MOLYB	1.97e+01
40	47	47.5	470	1	NRAM_IAMHM	NEURAMINIDASE (EC 3.2.	1.97e+01
41	47	47.5	512	1	AMY_BAGLI	ALPHA-AMYLASE PRECURSO	1.97e+01
42	47	47.5	544	1	TCPH_MOUSE	T-COMPLEX PROTEIN 1, E	1.97e+01
43	47	47.5	606	1	POUG_YEYF8	DNA TERMINAL PROTEIN (1.97e+01
44	47	47.5	1163	1	FABI_YEAST	GENOME POLYPROTEIN (CO	1.97e+01
45	47	47.5	2278	1	FABI_YEAST	PROBABLE PHOSPHATIDYL	1.97e+01

ALIGNMENTS

RESULT 1
 ID DSG3_HUMAN STANDARD: PRT: 999 AA.

AC P32926;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).

OS DSG3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RP [1]

RX MEDLINE: 92069753.

RA Amagai M., Klaus-Kovtun V., Stanley J.R.;

RT "Autoantibodies against a novel epithelial cadherin in pemphigus

RT vulgaris, a disease of cell adhesion.";

RL Cell 67:869-877(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE

CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND

CC CARCINOMAS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS

CC (POTENTIAL).

CC -1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: M76482; AAA60230.1; -.

DR PIR: A41088; IYHU03.

DR HSSP: P09803; 1EDH.

DR MK: 169615; -.

DR PROSITE: PS00232; CADHERIN: 3.

DR PFM: PF00028; cadherin; 4.

DR Cell adhesion; signal; Transmembrane; Cytoskeleton; Glycoprotein;

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

THIS PAGE BLANK (USPTO)